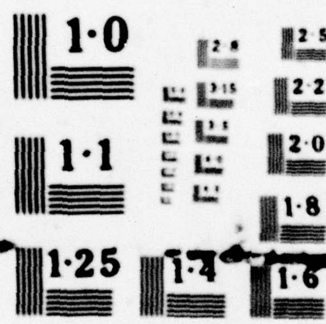
AD-A075 132 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 5/2
INTERNATIONAL ACCESS TO AEROSPACE INFORMATION.(U)
OCT 79

UNCLASSIFIED AGARD-CPP-279 NL

1 OF 1
AD-
A075132

END
DATE
FILMED
11-79
DDC



NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

LEVEL II

AGARD-CPP-279

AGARD-CPP-279

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AD A075132

AGARD CONFERENCE PREPRINT No. 279

International Access to Aerospace Information

D D C

RECEIVED

OCT 17 1979

A

DISTRIBUTION STATEMENT A

Approved for public release; Distribution Unlimited

DDC FILE COPY

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY ON BACK COVER

79 10 16 006

(14) AGARD-CPP-279

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

(11) Oct 79

(12) 49

(9)

AGARD Conference Preprint No.279

(6) INTERNATIONAL ACCESS TO AEROSPACE INFORMATION.

Accession For	
NTIS GRA&I	✓
DDC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or special
A	

This Preprint is published for the information of those attending the AGARD meeting at which these papers will be presented. Preprinted papers are subject to amendment by the authors before final publication in the AGARD Conference Proceedings series.

Advance copies of papers to be presented at the Technical Information Panel's Specialists' Meeting to be held in Athens, Greece 17–18 October 1979.

400 043

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

- Exchanging of scientific and technical information;
- Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;
- Improving the co-operation among member nations in aerospace research and development;
- Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;
- Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;
- Providing assistance to member nations for the purpose of increasing their scientific and technical potential;
- Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published October 1979

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

THEME

One of the main elements of the work of the AGARD Technical Information Panel (TIP) is to assist NATO's aerospace research and development activities by improving the effectiveness of scientific and technical information systems which exist in aerospace and defence fields in the various member nations of the Alliance. The choice of theme for this 1979 Specialists' Meeting stems from consideration of this aspect of TIP's work, particularly in relation to the host country, ~~namely~~ Greece. Development goes on constantly, and the Meeting will review recent changes not only in the channels and media of information transfer but also in the nature of the information itself. Arising out of this, it is hoped that firm plans can be laid for a broadening and extension of existing information transfer facilities in the fields of aerospace and defence R and D in particular areas of the NATO Community.

This document contains preprints of eleven papers to be delivered at the conference. Topics of the four sessions are

(cont on p. v)

TECHNICAL INFORMATION PANEL

CHAIRMAN: Ir A.S.T.Tan
National Aerospace
Laboratory (NLR)
P.O.Box 90502
1006 B.M.Amsterdam
The Netherlands

DEPUTY CHAIRMAN: Mr H.E.Sauter
Administrator
Defense Documentation Center
Cameron Station
Alexandria, VA 22314
USA

PROGRAMME AND MEETING OFFICIALS

PROGRAMME COMMITTEE CHAIRMAN: Ir A.S.T.Tan
(see address above)

PROGRAMME COMMITTEE MEMBERS:

Mr R.Bernhardt
Leiter der Hauptabteilung Datenverarbeitung
GID, Herriotstrasse 5
D-6000 Frankfurt am Main 71,
FRG

Mr D.W.Goode
Chief Librarian
Royal Aircraft Establishment
Procurement Executive, MOD
Farnborough, Hants GU14 6TD
UK

Ing. Gén. C.Guilleminet
Directeur Adjoint du CEDOCAR
26 Bd Victor
75996 Paris Armées,
France

Mr H.K.Krog
Managing Director
Norwegian Centre for Informatics
Forskningsveien 1
Oslo 3
Norway

Mr H.E.Sauter
(see address above)

MEETING COORDINATOR

Captain G.N.Goulios
KETA
Delta P Phalirou
Palaion Phaliron
Athens
Greece

PANEL EXECUTIVE

Mr E.T.Sharp

CONTENTS

* Not available at time of printing.

European Cooperation in the Field of Aerospace Information

W.A. Martin

European Space Agency

Information Retrieval Service

ESRIN, Via Galileo Galilei

00044 Frascati, Rome, Italy

SUMMARY

Agreements reached in the early sixties leading to the establishment of the ESRO/ELDO Space Documentation Service are reviewed. European cooperation in both the input of aerospace information and its utilisation are described. Details of ESA-IRS activities in support of the NASA information system are given. The development of the use of aerospace information in Europe as evidenced by the growth of IRS networking is examined. Research into aerospace information handling is outlined. The relationship between ESA and the CEC is explained in the context of EURONET.

1. LAYING THE BASIS FOR EUROPEAN COOPERATION : THE ROLE OF ESRO-SDS

1.1 In May 1964, an agreement for cooperation and exchange in the field of scientific and technical information was reached between NASA and the European Space Research Organisation. Early the following year, in April 1965, an agreement to set up a joint documentation service entered into force between ESRO and its fledgling sister organisation ELDO - the European Organisation for the Development and Construction of Space Vehicle Launchers. The result was the creation of SDS - the Space Documentation Service.

1.2 Under the terms of the agreement reached between ESRO and ELDO the joint documentation service was to cover both space research and space technology, and was to serve not only the staff of the two European space organisations but also their member states. The recipients in the member states were originally defined as "government establishments, universities, and institutes working in international or national space programmes", and "qualified persons or companies residing in the member states of ESRO and ELDO". The possibility of providing services to organisations in non-member states was also foreseen in a paragraph stipulating that such requests must be referred to the ESRO Council.

1.3 A further agreement was reached late in 1966 between ESRO/ELDO and EUROSPACE - the European Industrial Group for Aerospace Studies. This agreement enabled Eurospace member companies to obtain services from SDS, ie. bibliographic searches and/or selective dissemination notifications (SDI). It contained a financial guarantee to SDS for a minimum sum of 75 000 French Francs, and the requirement that "Eurospace shall endeavour to prevail upon its members to increase the number of their subscriptions and to have them renewed, and shall encourage the recipients to take advantage of the (Space Documentation) service beyond the facilities covered by their subscriptions". This preoccupation with financial r[illegible] was entirely understandable when it is recalled that not only SDS but also its parent organisations were in their very early stages of development. Thus from the beginning SDS has been required to provide services to all sectors of the aerospace community within the member states of ESRO and ELDO.

1.4 A significant aspect of the 1965 ESRO/ELDO agreement concerned the types and sources of material to be provided by ESRO and ELDO to the joint documentation service, and an acknowledgement that these could be made available to NASA. "ESRO and ELDO shall provide scientific and technical documentation, of which the publication is not prohibited, received or produced by them during the performance of the tasks described in their conventions or resulting from these tasks. They will in particular ensure that in research, study, development or other relevant contracts ...... the application of this provision is duly observed.

ESRO and ELDO agree that all documentation which is made available under the terms of this agreement may be channelled to ...... NASA ......; documentation received from NASA ...... shall be brought into the joint documentation service."

The original NASA/ESRO agreement of 1964 included the requirements that "ESRO will provide to NASA abstracts of scientific and technical reports originating from European sources ...... processed in a form suitable for inclusion in NASA's Scientific and Technical Aerospace Reports (STAR)", and "NASA and ESRO will make available to each other single copies of microforms ...... representing ...... abstracts published in STAR." In addition, it was stated that "NASA and ESRO have agreed, in principle, to exchange material for computer searches ......"

1.5 In looking back through those early documents dating back some 15 years, I was impressed by the very thorough job which seemed to have been done of setting the stage for a good two-way flow of aerospace information between SDS for Europe and NASA for the United States, and for its proper dissemination throughout the member states of ESRO and ELDO. It is clear that, a decade and a half ago, the manner in which Europe proposed to cooperate in the field of aerospace information had been defined in considerable detail. A tangible demonstration of its resolve to realise this cooperation lay in the creation of SDS, the Space Documentation Service, the role and activities of which had largely been specified as a result of agreements now dating back some fifteen years. The subsequent development of the Service lay in the manner and extent to which it was able to realise the mandate defined in principle in the mid-sixties. I will leave you to decide later the extent to which these early intentions have subsequently been achieved.

1.6 In 1973 the decision was taken to terminate the majority of the activities carried out by ELDO, and to re-orient those carried out by ESRO to include applications satellite projects in addition to the purely scientific missions previously flown as well as the development of a European launcher, ARIANE, and the re-usable Spacelab to be carried by the shuttle developed by NASA. With the fusion of the two original European space organisations a new agency named ESA, the European Space Agency, was created in 1975.

1.7 Later still, in 1978, ESRO-SDS, which had already become known as ESA-SDS, was given a new name. The intention was to reflect the wide range of scientific and technical information and services which were, by then, available from this Department of ESA responsible for STI and other information and data facilities. The new name chosen was IRS - Information Retrieval Service.

2. EUROPEAN COOPERATION TODAY, AS SEEN BY ESA-IRS

Today, IRS activities might be summarised as follows :

2.1 For NASA input, about 4 000 documents originating from European aerospace sources are handled annually of which almost 70 per cent are fully pre-processed to computer readable weekly tapes for STAR announcement. About 100 of these documents are selected for complete translation into English.

2.2 An online interactive search system is operated for ten hours daily on all working days. Some 15 major scientific and technical bibliographic databases plus two specially developed databanks total over 12 million records online at all times. Two databases each now exceed three million bibliographic references (Chemical Abstracts and PASCAL - Bulletin Signaletique), and are believed to be the largest online databases of their type in the world. Each database is updated monthly. Searching may be via a controlled vocabulary, if available, and/or via the natural language of title, title extension, abstract, uncontrolled terms or any other desired data element. A special numeric range search command has been developed for use with the databanks.

2.3 In parallel, an extensive STI network has been built up and provides direct online access to the IRS computer from all ESA member states and from several non-member states. Well over 10 000 kilometers of leased telephone lines stretch from Stockholm in the north, to Rabat, Morocco in the south. This network supports high-speed 2 400 bps (240 char/sec) video terminals equipped with 190 line/minute printers multi-dropped directly, and lower-speed 300 to 1 200 bps dial-up terminals via remote concentrators, provides remote offline printing via medium-speed printers and is used for various purposes by the Agency including facsimile transmission and computer to computer links. Other networks are interconnected to provide more access points and backup redundancy.

2.4 Today, more than 1 000 users of this system, with constant additions throughout currently thirteen European countries, one North American and one North African state, demand continuous support as facilities are improved and new databases added.

Recently an online order facility has been implemented which enables a user to order any original document announced in the NTIS and PASCAL files by means of a quick and simple operation at the terminal. There has been demand from both users and database suppliers alike to extend this feature to other files; as this paper is being prepared online ordering of NASA original documents is being implemented. For those users who require absolute security to surround their online file access a "TOP SECRET" command has been introduced which renders it quite impossible for anyone at the IRS computer centre to know what search the user is executing; no record is accessible other than by the user and this is destroyed instantly at logoff or if a break of any kind occurs. The more sophisticated terminals supported can be switched from retrieval mode to input mode (ODE) and can then be used to create machine readable data in a most convenient and efficient manner.

2.5 Some comments on the file coverage might be of interest. First, the needs of the Agency must be met as far as possible, ie. Agency staff, its advisory groups, its prime contractors, and its scientific collaborators. Because no multi-disciplinary mission oriented service can cover any single category within its scope to the same depth as a discipline-oriented service covering that category exclusively it has proved necessary to augment the NASA aerospace file with a number of discipline-oriented databases, particularly in the fields of physics, chemistry, electronics and engineering. More recently it was decided to add a biological database to cover the life science requirements of the SPACELAB programme; and an agricultural database, plus three small files covering oceanography, environmental and pollution information in anticipation of the information needs resulting from the new earth resources activities. For the latter a special databank of earth resources imagery, available from ESA as either photographic products or computer compatible tapes, has been developed and implemented.

The information needs of users in the member states tend inevitably to be much less well-defined than those of the Agency itself. The most marked demand was for chemical information which confirmed the need to add Chemical Abstracts Condensates. This file has become the largest IRS file and has the highest utilisation factor, over 22 per cent of total connect hours. In fact, chemistry, physics, engineering and general research together account for approximately two thirds of the total access time.

2.6 More recently there has been a growing demand from many member states for IRS to offer a more complete range of the information required. More specifically there has been criticism of the IRS policy of restricting the service to scientific and technical information and thus failing to satisfy the need for business and economic information. This leads to the question "just how can aerospace information be defined?", a question which has been asked many times but remains unanswered. One recent suggestion, for a very pragmatic approach, has been to carry out an analysis of which files are actually used significantly by the aerospace community. If this were done and the result was used as a basis for the definition of "aerospace information" it is clear that business files (which would enable searchers to obtain answers to questions of the type: "what large contracts have been obtained by company X recently?", "which company got contract Y in Saudi Arabia?", or "what are the turnover figures for the last 5 years for bidder Z?") would certainly be included.

2.7 As an indication of Europe's cooperative input to the world-wide aerospace information system operated by NASA, the IRS pre-processing statistics are incomplete due to the fact that the handover to IRS of all member states' document acquisition had not been completed during 1978. As a result, the numbers of documents sent from certain member states, notably the United Kingdom, did not represent the total contribution since some were sent direct to NASA.

As an indication of Europe's cooperative utilisation of aerospace information the IRS statistics of online access are also incomplete since they cannot take account of the searches performed manually by means of NASA STAR (Scientific and Technical Aerospace Reports) and IAA (International Aerospace Abstracts) which may still comprise a significant proportion of the total searches made.

3. EUROPEAN SUPPORT OF NASA SCIENTIFIC AND TECHNICAL INFORMATION ACTIVITIES : A CLOSER LOOK AT INPUT

3.1 IRS acts as a decentralised European extension to the NASA Scientific and Technical Information Facility. It maintains contact with all bilateral NASA partners in Europe and more recently the "tripartites", and arranges information exchange agreements between ESA and European groups working in the aerospace field. All this is intended to ensure the acquisition of as much as possible of the aerospace literature originating in Europe. An attempt is also made to cover conference schedules and programmes. All documents received by IRS are registered, and credited to the forwarding organisation. An Acquisitions Database is updated on a weekly basis and from this are generated half-yearly and yearly detailed and summary reports which are sent to NASA, to all ESA delegations, and are used by IRS itself.

3.2 After registry, all items are submitted to a selection procedure which can result in any of the following treatments :

3.2.1 IRS pre-processing to computer readable tape plus microfiche for announcement in NASA STAR

3.2.2 Despatch to AIAA for announcement in International Aerospace Abstracts

3.2.3 Pre-processing for initial STAR announcement; cover-to-cover translation; pre-processing for second STAR announcement of English version

3.2.4 Discarded as it duplicates an item already in the NASA Information System

3.2.5 Rejected or discarded as out-of-scope or otherwise unsuitable.

3.3 When an item is selected for STAR announcement it is given an English language abstract (the source document may be in any of eight European languages including English); descriptors from the NASA Thesaurus are assigned; it is catalogued; and a mother microfiche of the source document is prepared. Simultaneously, the details coded into the input worksheet (the form 901) are keyboarded to computer readable ATS compatible form using an online input facility. Weekly, the cumulation of data is rolled out onto a mini-tape and sent to the Facility with a back-up copy. Also from this ATS data is stripped the input to the IRS Acquisitions Database.

3.4 At the request of NASA certain series of reports issued by ONERA (Office National d'Etudes et de Recherches Aerospatiales), France and DFVLR (Deutsche Forschungs- und Versuchanstalt für Luft und Raumfahrt e.V.), Germany, are considered automatically to be candidates for full translation by IRS unless rejected for some reason, usually out-of-scope. More recently ONERA has offered to provide the periodical La Recherche Aerospatiale, issued six times a year, in the form of a cover-to-cover translation into English. Following translation the English language version of the report is issued in the ESA TT-series (Technical Translation), pre-processed for STAR, and cross-referenced to the original announcement of the source document in the original language.

3.5 A special report acquisition and STAR pre-processing project was initiated at the end of 1974 as a result of the close-down of ELDO. Due to differences in the Conventions of ESRO and ELDO the latter had not during its lifetime been able to make available a comprehensive collection of technical reports. As part of the liquidation programme almost 15 000 contractor reports and other documents generated during the ten year life of ELDO were examined, of which more than 1 600 documents, containing almost 2 000 discrete papers, were selected for full pre-processing for NASA STAR announcement. These items were all announced in L-STAR but, perhaps more important, now form part of the NASA computer readable database.

3.6 In addition to the largely passive role of acquisition of the European literature, IRS performs an active monitoring and liaison function with NASA's European bilateral partners and, more recently with the "tripartites". The latter organisations are those which wish to have direct online access to the NASA File via the Agency's STI network and online service. When an IRS user requests access to the NASA File he is asked to sign a tripartite arrangement between NASA, ESA and the user organisation. Under the terms of this tripartite arrangement NASA agrees to grant online access to the NASA File in return for the submission of literature within the scope of the NASA Information System. Items published openly, eg. in the periodical literature, are not accepted since these would be processed by AIAA independently of such an agreement. Difficult-to-obtain material such as research reports are sought. From NASA's viewpoint there is little difference between a bilateral and a tripartite. In each case an exchange is involved although in the case of tripartites no material dissemination is required.

3.7 The monitoring function is carried out on the basis of a number of system aids. First, the overall performance of each source organisation is checked annually on a "this year, last year" basis. Second, a regular report is generated from the Acquisitions Database which lists all sources which have sent zero input, one document, two and so on, over the prior period. The sources are listed alphabetically by country and where appropriate, contact is made with the source to draw attention to the absence of, or drop in, input and to establish and, if possible, correct the cause.

In addition to these guides, a special report on all tripartites is produced each half year. This report lists the number and type of documents submitted to IRS by each tripartite organisation and the number of connect-hours in the NASA File.

4. EUROPEAN UTILISATION OF AEROSPACE INFORMATION : IRS'S EXPERIENCE WITH NETWORKING

4.1 Towards the end of 1968, the 400 000 NASA File records needed 20 hours of computer residence time to process a batch of 30 questions; the operation would clearly soar out of hand in the following year. Following an evaluation of available systems and discussions between NASA STIO staff and SDS the decision was taken to implement RECON so as to give an inverted file system to shorten search times and increase capacity, and to have a system compatible with that used by NASA.

4.2 RECON was implemented on ESRO's largest IBM 360 computer installation which was located at ESOC (European Space Operations Centre), Darmstadt, near Frankfurt. In addition to the advantages already mentioned, RECON offered the bonus of being an online system and the original ESRO STI network consisted simply of two leased international telephone lines, one from Darmstadt to Paris where a RECON terminal was installed at ESRO Head Office, and from Darmstadt to Noordwijk near Amsterdam, to support a terminal located in the main library of ESTEC (European Space Technology Centre).

4.3 During that same year great interest was expressed by several member states in introducing direct online access to the SDS database in their national aerospace information centres and, with the approval of NASA, the first such "external" terminal was installed in 1970 at the Technology Reports Centre, near London. The embryo STI network was beginning to grow. The first installation in the UK was quickly followed by one at ZLDI (Zentralstalle für Luft- und Raumfahrt Dokumentation und Information), Munich, and by the end of 1975 some 20 high-speed leased line video terminals had been installed in the member states.

4.4 In 1973 SDS, which had been split between ESRO HO, in Paris, and ESOC, Darmstadt, was transferred to new premises at ESRIN, Frascati, near Rome, which had previously housed fundamental space science research activities. This brought all SDS staff together under one roof and provided the Service with a dedicated IBM 360/50 computer in support of STI activities for the first time.

4.5 To minimise the problems of relocation, it was decided not to disturb the network, then essentially a STAR network centered on Darmstadt and operating at 2400 bps. The intention was to retain the existing network topography and to link ESOC and ESRIN by means of two high-speed channels of 9600 bps using multiplexing modems on geographically independent routings. When the two trunks were handed over by the PTT's however, it did not take long to discover that the line routings were identical; since when one line went down, very often both went out together! This was disastrous for network operation, of course, since any outage on the ESRIN link killed the entire network. The alternative routing was never achieved since apparently it never proved possible to reach the line quality required for data transmission via the only other possible route over the Alps!

4.6 The only solution seemed to be a total revision of the network to bring the "centre" to Rome. There were other factors giving a push to this project such as an increasing user demand for the support of more sophisticated hardware, eg. dial-up concentrators and remote line printers. Following this reconfiguration total network blackouts are now rare.

4.7 In the original concept the network was required to support several tens of 2400 bps terminals multi-dropped from a STAR of lines operated at 2400 bps. As far back as 1973, however, SDS had foreseen the need to introduce much cheaper access to the network. The high-speed leased line terminal was a sophisticated and responsive device but, certainly in European information terms, it was expensive, and could really only be justified in very busy information departments or services. The development of the RTC, or Remote Terminal Concentrator, was disappointingly slow - the first version was not available for testing until 1975. The RTC is based on a PDP-11/10 minicomputer and as originally tested acted as a speed switch from the 2400 bps transmission from the central computer to the PDP-11 on one side, and from four to ten 300 bps dial-up devices on the other. Once again the Technology Reports Centre, near London, was the first member state organisation to install the new device, the penalty for which was a six-month teething period during which would-be dial-up users in the UK were called upon to exercise some patience! The RTC, which behaves in a manner very similar to a 2400 bps terminal, may be multi-dropped from the same 2400 bps line as a group of LLTs in more or less any combination. RTCs have been installed in Copenhagen, Dublin, London, Madrid and Stockholm, whilst concentrators using multiplexers are installed in Brussels, Darmstadt, Noordwijk and Paris. Direct dial-up to Frascati is also possible.

4.8 The introduction of concentrators into ESANET has enabled a new level of service to be provided to users who previously had to request a search to be executed for them by intermediaries in national centres. Though the latter were using an online system, to the user it looked like "batch". Now, through the RTC, he is able to search directly, online, himself. This has led to a change in the role of the intermediary, from that of searcher to that of teacher. It has also led to new demands for yet higher levels of service - RTC centres wanted the printout via their own line printers, driven over the network outside, and even during RECON time, thus making extra demands on the network.

4.9 Dial-up users in France are particularly well-served as a result of the decision to develop a gateway linking ESANET with CYCLADES. The latter is an experimental, but advanced, packet-switched network with nodes in Paris, Rennes, Grenoble, Toulouse, and Lyon. This gateway, developed by the French, permits one way access to the IRS computer. A new network named TRANSPAC is expected to replace CYCLADES in France this year and plans are in hand to connect IRS with TRANSPAC.

4.10 Since 1975, IRS has also been connected to the TYMSHARE network which has nodes located in Europe at Paris, Brussels, The Hague, Zürich and Frankfurt. Particularly useful to IRS users located in the Netherlands and Switzerland this network provides additional capacity to cover peak traffic times, and backup redundancy in the case of network failures.

4.11 More recently IRS has been connected to the Rome node of DARDO, the Italian link with the North American networking services including TYMNET and TELENET in the United States and TELEGLOBE in Canada.

4.12 As this paper is in draft, preparations are well advanced for the connection of IRS to EURONET; this is covered in the next section.

4.13 Overall utilisation of the IRS online service via the network complex described amounted to 22 938 hours in 1978.

5. EUROPEAN RESEARCH INTO AEROSPACE INFORMATION HANDLING : THE IRS INTEGRAL DATABASE CONCEPT

5.1 We have carried out a number of experimental studies using the NASA File in order to try to measure some of the RECON system parameters, particularly in 1972/73 (1) and more recently when we reported some unexpected findings to the 1977 Cranfield Conference (2). Test searches were carried out by experienced users in several member states, also by IRS and other ESA staff. In the most recent study the analysis of the results was carried out by Cleverdon. Broadly speaking the results of the measurements of recall and precision compared with those found by Lancaster and others in studying searching of the National Library of Medicine's MEDLINE system. Our earlier study suggested that in day-to-day use, on average, about 50 per cent recall at 68 per cent precision could be expected for an elapsed terminal time of about half an hour per search. The recent study did not try to measure absolute recall, but only comparative recall. However, correlation of these figures with the earlier study suggests very similar results of 51 per cent recall at 74 per cent precision with elapsed terminal time at 38 minutes.

5.2 Our experience with the NASA File and with several other major databases led us to believe that the NASA File is about the best example of what can be achieved using a well-controlled vocabulary. However, for any database, typical results of 50 or 60 per cent recall should not permit complacency since, for every 5 or 6 documents found which are known to be relevant to the question, a further 4 or 5 equally relevant documents remain "mislaid" in the computer.

5.3 Analysis of controlled retrieval language search failures in the experimental studies readily demonstrates the inconsistency with which valid descriptors from the controlled vocabulary are assigned to similar documents. In other words it is one thing to achieve a well-controlled vocabulary, but quite a different matter to control the indexing based on that vocabulary. We are probably unlikely to improve matters as long as manual procedures are used for checking. In my view, a well-structured thesaurus containing more than a very few thousand terms is far too complex an instrument to be used unaided by the human indexer.

5.4 There is also the need to make it easier for a user to search across several files, preferably simultaneously. However, the studies carried out had served only to emphasise the difficulties and failed to produce any realistic proposals to solve the problem - all approaches were based on the notion of cross-linking the various thesauri. Such a concept is operationally unattractive, smacking of the sledgehammer and nut tactic. Some indication of the complexity of such a task is given in a paper by Niehoff (4) describing the development of an integrated controlled vocabulary in the relatively limited area of energy terminology alone. In my view what was needed to achieve an integral database was a common retrieval vocabulary for all the files to be merged.

5.5 The idea emerged of the possibility of using the natural language of title, title extension, abstract, free-terms, etc, as the primary retrieval vehicle to be used with several files in an integral database and regarding the retrieval keys provided by the tape supplier (eg. classification code, descriptors, specialised codes such as CAS Registry Number, etc) as auxiliary search aids specific to individual files.

Though a great deal of work had been reported on the use of natural language for information retrieval it was not possible to draw any hard and fast conclusions. Most work had been done offline and it was difficult to compare results reported due to the surprising variety of bases for the experiments.

Opinions tended to polarise strongly in one direction or the other, often it seemed, without adequate justification. Perhaps the following quotes will illustrate this situation :

- "Experiments already carried out have demonstrated that natural language, with minimal or no control, is superior to any form of controlled vocabulary." (5)

- "The reasons for the failure of free-text indexing are both theoretical and practical. Many computer experts continue to propose the computer implementation of free-text indexing either ignorant of, or indifferent to, past failure." (3)

5.6 We began to feel that if retrieval performance based on natural language searching could be developed to the point where it was not significantly inferior to that achieved when using a controlled vacabulary, then it could provide the vehicle for an integral database. An experiment was designed by IRS, a test database was created online based on the NASA STAR tapes for 1973 and 1974 with the agreement of NASA, and an evaluation was carried out under the direction of Cyril Cleverdon (6). Experienced searchers from several European countries carried out test searches using either controlled vocabulary or a special natural language index. This experiment was reported and the results discussed at the 1977 Cranfield Conference (2).

5.7 The results surprised us and could be summed up in the following quote from the cautious Cleverdon :

"It appears difficult to reach any other conclusion than that, within the parameters of this test, natural language searching on titles and abstracts proved at least equal to and probably superior to searching on controlled language".

In fact for virtually every parameter used to compare the two techniques of searching, ie. natural language and controlled language, the former was demonstrated to be superior.

5.8 In my own view, these results have shown that retrieval using the developed natural language index was not only "not significantly inferior to" but was actually superior to that obtained on the controlled vocabulary. The techniques needed to create the index were tested and shown to be perfectly practical. We can see that retrieval on titles only is futile though it could be a convenient adjunct to a controlled vocabulary search. However, the amount of processing time needed to create the inversion on the abstract text is not insignificant, whilst the disk demand for the inverted files can more than double, when compared with the basic IF on descriptors, authors and corporate source.

5.9 I hope that further work will be possible on the integral database concept. I believe the results obtained to date, with an index which incorporated only one of three planned approaches and is certainly capable of further improvement, have already demonstrated the feasibility of the approach. I hope also to see the experimental NASA STAR File further developed according to the original plans and to see a second database given identical processing and the inverted files merged. This would constitute the first experimental integral database for evaluation.

5.10 At a different level complete re-design of the online applications software has been in progress for over two years and a very new version is on the way which will include the implementation of the "EURONET Common Command Set" referred to later.

5.11 A policy of database standardisation has been developed such that the different files should appear to be more homogeneous to the user. A series of standard formats has been specified for display, type and print outputs. As these are implemented on all the files, it means that any given data element, be it a document type or a classification code, will always appear at the same position in the record; data elements which are identical but are described by different terminology by the database producer (eg. free-term, identifier, keyword, etc) will be given standard names, eg. controlled term (for descriptor in NASA parlance); uncontrolled term (for identifier on COMPENDEX); category code (section number in Chemical Abstracts). Where possible, abbreviations will be expanded to the full form and a standard table of language names will be introduced. Prefix codes (AU=) and Suffix codes (/CS) in the inverted file will be standardised.

6. EURONET : THE RELATIONSHIP BETWEEN ESA AND THE COMMISSION OF THE EUROPEAN COMMUNITIES (CEC)

6.1 The name "EURONET" means many things to many people and indeed, embraces a very wide spectrum of STI possibilities. A significant amount of money has been allocated to the project - approximately 6 M dollars for the first 3 year period 1975-1977, and a similar amount for the second 3 year period 1978-1980. One CEC spokesman has explained it by saying : "We wish to create a Common Market in scientific and technical information ...... information is a commodity which can be bought and sold like any other".

6.2 EURONET will be a packet-switched STI network serving the member countries of the CEC (France, Italy, Germany, Belgium, Luxembourg, the Netherlands, Ireland, Denmark and the United Kingdom). It will also encompass Community policies on STI, embracing the types of database and databank needed in the CEC countries and including the sponsoring of new databases and databanks where these are seen to be required, and on multi-lingualism ultimately intended to permit access to computer data via all Community languages.

6.3 EURONET as a physical network is being created by a consortium of all nine PTT's from the CEC member countries. The current target date for operation is autumn 1979. The presently foreseen network will be a two-level system with four network switching nodes located at Frankfurt, London, Paris and Rome and five remote concentrators at Amsterdam, Brussels, Copenhagen, Dublin and Luxembourg. Initially, only teletype compatible terminals will be supported. The first host computers scheduled to be connected to EURONET include those of IRS; DIMDI (Deutsches Institut für medizinische Dokumentation und Information) in Cologne; INKA (Fachinformationszentrum Energie, Physik, Mathematik) in Karlsruhe; the British Library Service BLAISE; and a second UK venture named INFOLINE.

6.4 Of the nine member states of the CEC and the eleven of ESA, eight are common to both. It is perhaps hardly surprising that, in view of the comprehensive European STI capability already developed by IRS, and the far-reaching plans of EURONET, ESA and the CEC should be encouraged by the member states to discuss opportunities for cooperation. The result was an exchange of letters between the Director General of ESA and the Director General of the Directorate General for Scientific and Technical Information and Information Management of the CEC.

6.5 Europe needs a reliable trans-European TYMNET type STI network as soon as possible. To help achieve this, following the exchange of letters ESA has collaborated with the CEC to the maximum extent possible. This has included the detachment of two IRS staff to the CEC's DG XIII in addition to a very considerable amount of staff time which has been allocated. A so-called "Mixed Team", comprising three CEC staff and three from ESA, all at high level, is charged with the practical implementation of the cooperation agreed. In addition, IRS staff are inevitably involved in many committee and working group activities associated with EURONET.

6.6 Although EURONET suffered the inevitable delays of any large-scale multinational project, by the end of 1978 there were unmistakeable signs that it would be available in the comparatively near future. A contract was placed with a large marketing agency for the pre-launch phase; a host-network technical unit was created to coordinate host connections; and a user relations unit was set up; a launch team was being established to "help the smooth running of EURONET operations". EURONET also acquired a new name - DIANE ! The new acronym stands for "Direct Information Access Network for Europe".

6.7 By March 1979 the CEC announced that the REX-25 testing facility which had been opened to host information services was being used successfully; this facility (located in Rennes, France) simulated the X-25 standard interface to be used for host-to-network links on EURONET. At the same time it was announced that switching equipment had been delivered to all sites with the exception of Rome! Negotiations with Switzerland and Spain for EURONET access were proceeding well and a dialogue had been opened with Sweden.

6.8 IRS' own preparations for connection with EURONET by mid-1979 were extensive. Delivery of the "black box" device to be used for the connection, a TERPAC H-12A unit made by the French company Sitintel, was scheduled for end-June, to coincide with the installation by SIP (the Italian telephone company) of the line to connect IRS with the Rome node of EURONET. About two months of testing were foreseen, falling in the seasonally quieter months of July and August, leading to an operational EURONET connection by September. Major changes had been under development for the applications software for some considerable time, including the provision of a totally new command language, CCS - the EURONET Common Command Set specified by the CEC for all EURONET hosts. The target for September 1979 was release of version 1 of IRS' implementation of CCS, with the user being offered the choice of IRS' traditional command language or the new CCS by means of a switch.

REFERENCES

1. Martin, W.A. A comparative study of terminal techniques ...... Paper presented to the First European Congress on Documentation Systems and Networks, Luxembourg, 16-18 May 1973, EUR 5058, pp. 107-164

2. Martin, W.A. Publication pending

3. Wessel, A.E. Computer-Aided Information Retrieval. Melville Publishing Co., Los Angeles, Calif., 1975, 176 p. ISBN/ 0-471-93376-7

4. Niehoff, R.T. Development of an integrated energy vocabulary and the possibilities for online subject switching. J. American Soc. Inf. Science, V27, n 1, Jan - Feb 1976, pp. 3-17

5. Bhattacharyya, K. The effectiveness of natural language in science indexing and retrieval. J. Documentation, V30, n 3, Sept., 1974, pp. 235-254

6. Cleverdon, C.W. A comparative evaluation of searching by controlled language and natural language in an experimental NASA database

The Role of NASA for Aerospace Information

George P. Chandler, Jr.
Chief, Scientific and Technical Information Branch
National Aeronautics and Space Administration

INTRODUCTION

In 1962 the United States announced its commitment to put a man on the moon by the end of the decade. In the exciting and stimulating environment created by the launching of this commitment NASA began its activities for providing scientific and technical information to the scientists and engineers working to solve the problems of manned space flights. This year we observe the tenth anniversary of man's flight to the moon and safe return to Earth, -by any measure a remarkable feat of applied engineering and of human technical genius.

But from the beginning NASA recognized that its goal of putting a man on the moon was but one part of an aerospace program of research and development that would provide a broad variety of benefits to all mankind. Stated clearly among NASA's goals is the need to disseminate as widely as practicable the scientific and technical knowledge developed as part of the nation's aerospace activities. Today the NASA STI program has become a valuable national and international resource with more than 1.6 million documents recorded in its data base and growing at a rate of 80,000 documents per year. The collection of documents continues to grow in subject scope as well as in size in order to meet the needs of new NASA programs, aerospace programs of other nations and a growing diversity of users. In responding to the needs of this growing universe of users, NASA has developed a highly sophisticated and effective automated system of providing direct access to its data base. It is a dynamic and flexible system that constantly monitors requirements for information by government, industrial and economic groups throughout the world. The needs are translated into new services and products many of which are made possible only through the effective utilization of new techniques and new technologies.

Despite the reduced level of the current U.S. space program, because of budget limitations, many exciting programs are underway: the Space Shuttle, planetary probes, satellite surveys, and others. In addition, many more nations throughout the world now have ongoing programs of research and development related to aerospace activities that create a continuing flow of new information. Certainly an important current need reflected in the incoming flow of documents and requests for information is the design and construction of aircraft to utilize fuel supplies more effectively. Thus, an increasing audience of users continues to require information on space related developments as well as on the utilization of space-developed technology. Today, the system serves more than 11,000 scientists and engineers within NASA, some 100 NASA contractors, subcontractors, and grantee organizations, and more than 600 elements in U.S. Government agencies. Some 60 countries and over 750 governmental organizations, universities, and research institutes are active participants in the NASA International Exchange program. Through the European Space Agency, much of the NASA data base is made available across Western Europe. Increasingly, the public and commercial interests not directly engaged in aerospace activity are making use of the system.

On a typical day the NASA Scientific and Technical Information Facility receives requests for information from a wide spectrum of users. The president of a manufacturing company asks for information relevant to a materials problem. A nine-year old child writes to ask for information on how to build a rocket. In addition, each day hundreds of scientists, engineers, administrators, legislators, and others access online the NASA Scientific and Technical Information data base. They seek information that will help solve problems not only in the fields of aerospace activity, but also in such critical areas as the increase of the world's food supply, control of floods, location of new sources of minerals and energy, and in providing accurate warning of impending natural disasters.

Scope and Coverage of the NASA System

To achieve completeness of coverage and quality of content, NASA has an aggressive policy of seeking sources of documents. NASA utilizes an established worldwide network of cooperative public and private sources, from which it collects each year about 90,000 - 100,000 documents. These incoming reports as well as journal articles, reviews and books from government, industry, research institutions and universities contain the findings of physical scientists, social scientists, medical scientists, engineers and others who contribute to and are concerned with progress in aerospace development. Along with these reports and published articles, the system regularly receives documents and information on NASA-owned inventions covered by U.S. patents and applications for patents, contracts let by NASA, NASA research and technology programs underway or planned, and related research in progress. Such information is also gathered from other U.S. Government agencies and other major sources of aerospace information throughout the world.

The subject matter of these documents ranges widely and includes such areas as aeronautics, astronautics, chemistry and materials, engineering, geosciences, life sciences, mathematical and computer sciences, physics, social sciences, and space sciences.

The NASA STI system is managed by the Scientific and Technical Information Branch (STIB) at NASA Headquarters in Washington, D.C. The actual operations are primarily performed in two contractor-operated facilities. The NASA STI Facility, located near Baltimore, Maryland, employs about 210 people

who process report literature, operate the computer complex and provide support for software maintenance and developments. A second contractor, the Technical Information Services of the American Institute of Aeronautics and Astronautics, employs approximately 80 people in New York City and processes the open literature -- journals, magazines, books, etc.

The scope and versatility of the NASA STI program are apparent from the following service highlights:

1. Publication of four announcement journals, two on a biweekly basis, one on a quarterly basis, and one on a monthly basis.

2. Publication of a biweekly current awareness service covering almost 200 separate information categories.

3. Publication of six continuing bibliographies covering specific fields such as "Energy," "Earth Resources," and "Aeronautical Engineering."

4. Distribution of about three million microfiche per year.

5. Distribution of approximately 1.5 million hardcopy documents per year.

6. Provision of interactive data bank and retrieval capability to NASA Field Centers, NASA Industrial Application Centers (IACs), and others having NASA/RECON (REmote CONsole) terminals.

7. Provision of documents and microfiche for sale to the general public.

8. Preparation of indexes to NASA Congressional Hearings, speeches, press releases, and management documents.

9. Provision of magnetic tapes and microfiche to ESA for service to the European aerospace community.

10. Provision to accept system input from ESA through machine readable magnetic tapes.

11. Operation of the NASA library network (NALNET), which provides online search and retrieval access to all books held by NASA libraries.

Online Access Via RECON

The features of the NASA system which are most highly regarded by users are the NASA data base and the RECON system for online interactive access. Terminals link NASA installations and NASA Industrial Applications Centers throughout the United States to the Facility's central computer. Access is also available to other government agencies and to NASA contractors via dial-up capability. In addition to direct online searches via a terminal, data base searches are provided to registered users on request to the Facility in Baltimore, NASA Field Centers, and the NASA Industrial Application Centers. More than 11,600 searches are being made each year (3,600 at STI Facility; 8,000 outside the Facility) and the usage continues to grow. Currently, NASA Centers and Industrial Applications Centers have started using new intelligent terminals as the first step toward a distributed processing network that will provide many added software options at the local level as well as relieve the central computer of many routine tasks. Adjuncts to this retrieval system permit access to the data bases of the U.S. Department of Defense and the U.S. Department of Energy.

As we gain more and more experience in making computer searches with RECON, we become more impressed with the power and versatility of this interactive search and retrieval tool. Not only does it provide expanded capabilities and ease of use in meeting the information needs of scientists and engineers, but it is enabling others to use the scientific and technical literature as a source of economic and marketing data. For example, the results of such searches can produce information on the names of companies doing work in a particular field and thus suggest opportunities for new markets or for mergers or joint ventures. Information from reports of research and technology underway or planned is being used to identify newly developing fields which will foster tomorrow's markets and new technological advances. Increasingly, the NASA STI base is used to produce information of value to workers in the fields of contamination control, medicine, food, technology, and many others.

Since the computer provides only bibliographic citations and abstracts, it is essential that users be provided access to full documents as rapidly as possible. Regular delivery of documents, either in full-size printed form or in microfiche, is made by NASA through its STI Facility to more than 1,900 organizations in the U.S. and to more than 750 abroad. Regular delivery of documents and special publications is supplemented by sale through the U.S. Government Printing Office (GPO) and the U.S. National Technical Information Service (NTIS) to assure the widest possible public availability. Currently, a system of direct ordering of documents via the computer system is being tested.

A significant part of the computer software development program is concentrated on expanding the usefulness of the RECON system. Expanded usage creates demands for new capabilities and for greater ease in the use of the system. Each year, within the limits of our budget resources, we do as much as we can to incorporate new changes and to take advantage of new technology. In the past year, we have made significant progress in improved system reliability and addition of several important new features for users.

Currently being tested as a new capability of the NASA RECON system is a selective dissemination feature whereby individuals may establish and retain in the computer a profile of their specialized interests. This profile is then used to retrieve information covering a given time period, usually that time which includes all document accessions during the previous month and contained in the various

announcement journals. Users may update or change their profiles via a direct access terminal online. While intended to supplement existing alerting services, it is not unlikely that this type of computer capability may replace some printed products.

Announcement Services

The basis of NASA information services is a series of announcement journals generated from the computer based information files in the form of comprehensive or specialized bibliographies and abstracts varying in frequency of issue from semimonthly to quarterly. These announcement journals include: Scientific and Technical Aerospace Reports (STAR) which provides bibliographic citations and abstracts on some 1,000 accessions on report literature per issue; Limited Scientific and Technical Aerospace Reports (LSTAR) which provides abstracts and indexes of security-classified and limited distribution documents; International Aerospace Abstracts (IAA) which provides coverage of the published literature of approximately 1,600 accessions per issue. A service of particular value to individuals is Selected Current Aerospace Notices (NASA/SCAN). This biweekly service divides the content of STAR and IAA into some 200 subject profiles, thus providing highly specialized notification on new information entered into the system. STAR, IAA, SCAN, LSTAR, and other NASA announcement publications such as Computer Program Abstracts (CPA) serve to keep users aware of new additions to the document and information collection and to facilitate retrieval of required material. In all, about 40,000 journal articles and books and approximately 24,000 technical reports are announced each year.

To provide for prompt and effective dissemination of these publications, NASA maintains a computer-based mailing list for the generation of mailing labels and control of distribution. The considerable in-depth computerized data on users, products and services allows us to segment, manipulate and cross-match records so as to put any user -- or groups of users -- into categories which will permit us to examine statistically the who-where-what and how many of our audience and its demands. This computerized system assists in establishing resource requirements and allows evaluative analyses to be performed in setting future courses of action.

An important element of the NASA mission has become the transfer of aerospace technology for problem-solving and application in the private sector. This includes dissemination of information and online access to the NASA Data Base and also technical assistance in the evaluation of the commercial potential of a given product. The NASA STI Facility also maintains a computerized mailing list of some 20,000 organizations and individuals not directly connected with the U.S. space program who are mailed information about commercially useful spin-off developments from NASA research efforts. NASA stimulates interest in such opportunities through an active program of announcing innovations deemed to be of value to manufacturers for the development of new products. The principal such vehicle is a quarterly publication called NASA Tech Briefs. In further support of this effort, NASA has set up a technology utilization network currently comprised of seven regional Industrial Applications Centers which perform a type of "push or stimulation" dissemination. The IACs interpret the essence of specific available information for practical and useful applications and suggest its usefulness through technical consultation with potential developers. More recently, NASA has established two state technology centers to test the transfer of technology at the state level in the U.S. A major source of information for the IACs and the state centers is the NASA STI system.

International Document Exchange

In the early 1960s, prior to the NASA/ESA arrangement, NASA initiated and implemented an international document exchange program as one of its first moves to develop cooperative programs with European and worldwide organizations. The major objective of the program was to promote international cooperation through an orderly information transfer process. Governments, academia and selected research establishments within the respective countries were made exchange partners. NASA provided selected services, but primarily the semimonthly abstract journal STAR and its semiannual cumulative indexes in exchange for recent technical documents, journal articles and serials. As mentioned earlier, some 60 countries and over 750 governmental organizations, universities and research institutes are active participants in the NASA international exchange program. Receipts from this program and the ESA arrangement account for about 15% of the accessions in STAR and about 5% in IAA.

Recently a new concept was inaugurated in Europe to enhance the acquisition of documents for the NASA data base in exchange for access to the NASA STAR and IAA files via ESA RECON. It is our "Tripartite" agreement program. It stems from an agreement among NASA, ESA and a third party, e.g., an organization in a participating ESA member state. This program has enabled ESRIN-IRS to offer throughout the ESA community, via the ESA RECON network, direct online access to NASA STAR and IAA files. A major provision of the agreement is the requirement to provide one relevant document input to ESA (and thus NASA) for each hour of access to NASA files. At the moment there are about 300 Tripartite agreements throughout Western Europe. Results of the program and status of the Tripartites are reviewed semiannually.

ESA and NASA RECON

The mechanized, interactive online retrieval system at the European Space Agency is a result of a mutually beneficial exchange agreement between NASA and ESA (then ESRO) in 1962. At that time, NASA agreed to make available to ESA its monthly input of STAR and IAA references in machine-readable form in exchange. Today this input by magnetic tape to ESA now totals about 1,000,000 accessions.

In exchange, ESA through its ESRIN-IRS (Information Retrieval Service) at Frascati, Italy, supports NASA by providing the major document acquisition effort in Western Europe for STAR; providing a portion of the document acquisition effort for IAA; processing (cataloging, abstracting and indexing) on tape, in machine-readable form for direct input into the NASA system, about 2,200 STAR accessions and placing these accessioned items on microfiche; forwarding to the NASA STI Facility up to 1,500 other items for input into the NASA data file under one of the NASA unpublished accession series; producing about 100 English language translations of selected technical French and German language reports for input to the NASA STAR file; acquiring and providing to NASA select, difficult to acquire, Russian language aerospace reports through a special document exchange between ESA and the Institute of Space Research, Academy of Sciences,

Moscow. The Russian material is provided in exchange for ESA sponsored material. Microfiche masters are forwarded to NASA, along with the documents and their related processing forms, as well as as a machine-readable tape. NASA accession numbers are automatically assigned to the documents after they are processed for entry into the NASA data base. Periodical or open literature items are submitted for review to the Technical Information Service of the American Institute of Aeronautics and Astronautics and those selected for continuous receipt are acquired for inclusion in the semimonthly abstract journal International Aerospace Abstracts.

The online retrieval systems of NASA and ESA have a common ancestry in that both were initially developed by Lockheed in the late 60's and early 70's. Both RECON systems have since been considerably enhanced, and an effective exchange of information on new RECON developments is maintained between NASA and ESA. The ESA main computer facility is located at Frascati, Italy and serves the ESA RECON Network across Western Europe.

In 1975, ESA took on increased responsibility for the European area by collecting and processing the reports involving several hundred organizations. In 1977, NASA in cooperation with ESA produced the camera-ready copy of ESA SP-1006, INDEX OF ELDO PUBLICATIONS. Containing about 1,600 citations and abstracts with six computer-produced indexes, the document covers significant reports generated by and developed for ELDO, the European Launcher Development Organization.

NASA and AGARD

NASA engineers, scientists and technical information specialists have participated in the preparation of AGARD technical reports, books, manuals and journal articles since NASA became an active participant in AGARD. NASA STIB began more direct participation in the AGARD publications program in 1974 when the camera-ready copy of the AGARD INDEX OF PUBLICATIONS, 1971-1973, was prepared at the NASA Scientific and Technical Information Facility. This publication included an abstract section and five computer-generated indexes -- all from the NASA data base. Subsequently, another index for the period 1974-1976 was prepared. In addition, manuscript copies of the bibliographic sections for several AGARD Lecture Series topics, e.g., Strapdown Inertial Systems, Energy Conservation in Aircraft Propulsion, and Methodology for Control of Life Cycle Costs of Avionic Systems have been prepared by the NASA STI Facility. At the moment the AGARD Multilingual Aeronautical Dictionary is in production, and plans are being completed for the production of a third triennial index of AGARD publications for the period 1977-1979. The dictionary is a large and complex publication involving nine languages, including a need for many special characters and accents such as cyrillic and Greek. Scheduled for actual publication in early 1980, the dictionary will be a product of utilizing, at the NASA STI Facility, the latest developments in computer assisted composition and photocomposition technology.

The growing effectiveness and use of the NASA information system reflects the scope and the growth of aerospace programs by the nations of the world. The scope covers not only man's achievements in space, but the wide array of benefits accruing to man as a result of the research and development supporting those achievements. Increasingly, through the work of organizations such as the European Space Agency, those benefits are being made available more widely throughout the world.

We have attempted to use as much new technology, know-how and experience as possible in the building and operation of the NASA Scientific and Technical Information Program. We plan a continued program of keeping the system a highly effective source of information for all concerned with aerospace development and progress. In the near future we see the demands for access to the information contained in the NASA STI system more and more related to the rapid social, economic and political changes taking place throughout the world. It has never been more important that our information systems must be guided by a philosophy of readiness to change rapidly and effectively in anticipation of new needs and new demands. Not only for NASA, but for all information systems, the goals must be an increasing understanding of the ever more complex needs of our users, the provision of fuller access to all kinds of information, and the fulfillment of needs for specialized user requirements.

July 1979

STATE-OF-THE-ART OF STANDARDIZATION AND HARMONIZATION OF BIBLIOGRAPHIC DATA ELEMENTS

by

Harold Dierickx

The British Library, Research and Development Department (UNIBID)
Sheraton House, Great Chapel Street, London W1V 4BH, UK.

SUMMARY

As the title suggests, an attempt is made in the present paper to summarize the current situation with regard to standardization and "harmonization" of data elements in bibliographic descriptions. This is done with due regard to the increasing trend towards computerization of bibliographic information exchanges. Because of the growth in computer-aided processing of bibliographic records, the problem of choosing the appropriate headings for records has become less. Consequently, the emphasis in this paper is on standardization of data elements rather than on that of cataloguing codes. Nevertheless all aspects of the standardization of bibliographic description are considered, i.e. headings, bibliographic description proper, additional information. Only international and national rules and standards of international significance are dealt with.

Among the standards or guidelines which deal with all areas of the bibliographic record may be mentioned the "Anglo American Cataloguing Rules" (AACR), the "Regeln fur die Alphabetische Katalogisierung" (RAK) and uniform cataloguing rules drawn up by the USSR Cataloguing Committee for use throughout the Soviet Union. These may be considered as more library-oriented rules. Cataloguing rules which have been conceived more specifically for abstracting and indexing (A&I) and other secondary information services, which may or may not involve any library operations, are the so-called "COSATI Rules" or "COSATI Standard for Descriptive Cataloguing of Government Scientific and Technical Reports", the "American National Standard for bibliographic references" (ANSI Z39.29-1977), the "UNISIST Reference Manual for Machine-Readable Bibliographic Descriptions" (RM), "ISO 690-1975: Documentation - Bibliographical References - Essential and supplementary elements". Because of their general approach, two interrelated manuals developed in the Federal Republic of Germany for the cataloguing of bibliographic records processed by secondary information services, should also be mentioned in the present context: "Leitfaden fuer die formale Erfassung von Dokumenten in der Literaturdokumentation" and "Magnetband-Austauschformat fuer Dokumentationszwecke" (MADOK). All these are general rules intended for use by a variety of libraries and information services. They have often served as source material for rules designed for specific information services, such as the "International Information System for the Agricultural Sciences and Technology" (AGRIS) and the "International Nuclear Information System" (INIS). ISO 690 is more specifically designed for bibliographic citations in scientific publications. The A&I oriented rules put the emphasis on bibliographic description, assuming that headings will be created by individual systems on the basis of computer processing of separately identifiable individual data elements in bibliographic records.

The major source for creating headings to bibliographic records are the "Statement of Principles adopted at the International Conference on Cataloguing Principles, Paris, October, 1961". The incidence of computerization on the creation of headings has already been referred to. Other factors to be considered are the proposals to establish "universal bibliographic control" (UBC) via the creation of authoritative bibliographic records, including choice and form of headings, by national bibliographic agencies. Among other things, UBC might involve the generation and use of machine-readable international authority files to control the form of headings, e.g. authoritative form(s) of the names of persons and organisations. It remains to be seen how effectively UBC can be organized, especially from the point of view of the timeliness required by information services other than national libraries. It is unlikely that national bibliographic agencies are in a position to control all scientific and other literature to the detail and speed required by the different user communities. The constraints on UBC as originally conceived by IFLA are touched upon in the paper, mainly from the point of view of standardization of data elements.

As to standards for bibliographic description per se, an "International Standard Bibliographic Description" has been developed for most kind of bibliographic material (monographs, serials, non-book materials, cartographic material, etc.) under IFLA auspices. In addition to these ISBDs for specific materials, an ISBD (General) has also been formulated as a "harmonizing standard for all other ISBDs". The latest ISBD which is currently being formulated is that for analytical entries: ISBD (AN). Because of the necessity to work out a detailed and consistent definition and treatment of bibliographic levels for ISBD(AN), the treatment of bibliographic levels may have to be expanded in all other ISBDs where it is little developed. The different philosophies underlying general cataloguing rules and the ISBDs respectively are indicated.

Two of the more obvious examples of existing standards to convey information about bibliographic records which is not absolutely necessary for their identification, but which is nevertheless very useful, are "International Standard Serial" and "Book Numbers" (ISSN and ISBN). Report and patent numbering schemes as well as the proposed international sound recording numbering system are other examples. Internationally used rules for subject indication and filing, and the current attempt to standardise definition and use of "implementation codes" (status, bibliographic level, type of record etc.) are also discussed.

Although the [illegible] for standardization efforts to suit the different requirements of various user

groups is recognized, the conclusion reached is that standardization of form and presentation of bibliographic data elements (as distinct from introducing a universal and comprehensive cataloguing code) is an essential prerequisite to achieve international compatibility of bibliographic records, particularly in an automated environment. Finally, the point is made that the division between "libraries" and "all other information services", which has bedevilled so many standardization efforts in the past, is to a large extent artificial. The need for close and continued co-operation among the international (professional) organizations involved, in close consultation with their national member institutions, is stressed.

STATE-OF-THE-ART OF DATA EXCHANGE: PROBLEMS OF FORMATS AND STANDARDS

Dr. J.S. Mackenzie Owen
Netherlands School of Business
Nijenrode
3621 BG BREUKELEN
The Netherlands

SUMMARY - The vast growth of bibliographic data services over recent years has produced many problems in the field of compatibility for data exchange. It is now beginning to be understood that there exists a major gap between the library community on the one side, and the Abstracting and Indexing services on the other side. It is also becoming clear that any standardized format for bibliographic data exchange should be flexible enough to incorporate the divers needs of many different users. An effort is now being undertaken bu UNESCO to solve these problems. This paper describes the present situation, problems and future developments in this field.

BIBLIOGRAPHIC DATA EXCHANGE IN PERSPECTIVE

In historical times, the universe of documents available to any individual was more or less confined to the library or libraries in the city where that individual happened to live and work. Nowadays, the universe of documents available to an individual consists of almost all existing documents. This is not because libraries are nowadays so large as to contain all those documents, but because librarians have invented bibliographic data exchange and have integrated it with many and advanced modes of communication. Written and printed bibliographies were the first means of communicating to other people information about documents that were not immediately at hand. Bibliographic data exchange has also become a means of increasing efficiency. Documentation pools and distributed catalog systems have releaved libraries and information services of having to compile the same bibliographic information many times at different locations.

The huge problems created by the vast growth of scientific and technological publication have been solved to a great extent through the use of computers. And librarians were very early at doing so. This means that nowadays bibliographic information has to be exchanged in a form in which it can be understood and processed by computers. Such forms are commonly well defined information structures referred to by the term bibliographic formats. I shall return to bibliographic formats in the next sections of this paper.

The development of computers and related technology during recent years has resulted in on-line accessability of large-scale information systems as well as in advanced modes of communication. Individuals can now communicate with computerized data-bases on the other side of the ocean. Here, exchange of bibliographic data is a trivial matter. As little or no processing is reqired at the receiving end, it comes down to nothing more reading or printing out information in the data-base. This procedure is becoming more and more common, and, through systems such as Teletext and Viewdata may even bring bibliographic information into the private home. Very large bibliographic information systems and sophisticated communication networks may eventually supersede local processing of bibliographic data. In that case, there will be very little need for bibliographic data exchange in the traditional sense. The problem of standardization in man-machine bibliographic systems will move from the input side of the system to the output side, in the fields of access, retrieval and display formats. In my view, this will have detrimental effects on the flexibility of bibliographic information. Local processing means that the system can be adapted to local requirements, to the local document collection and the local type of user. Centralization and standardization are on hostile terms with individuality. The increasing collectivism and centralized control in society as a whole is, unfortunately, sure to be reflected in our future bibliographic information systems.

THE STRUCTURE OF BIBLIOGRAPHIC FORMATS

At the present stage of technological development, bibliographic data is usually exchanged on magnetic tape. Information on the tape has to have a well defined structure in order that the computer may identify individual bibliographic descriptions and their elements. A bibliographic description as a whole is referred to as a record. The exact lay-out of such a record on tape is defined by a well-known standard called ISO 2709, Format for Bibliographic Information Interchange on Magnetic Tape (1). ISO is the International Standards Organization, which has a special technical committee, TC46, for standardization in the field of libraries and documentation. The scope of ISO 2709 is described as follows:

> "This International Standard specifies the reqirements for a generalized exchange format which will hold records describing all forms of material capable of bibliographic description as well as related records such as authority records. It does not define the length or the content of individual records and does not assign any meaning to tags, indicators or identifiers; these specifications are the functions of an implementation format.
> This standard describes a generalized structure, a framework designed specifically for communications between data processing systems and not for use as a processing format within systems. Although this International Standard is designed for magnetic tape, its structure may be used for other data carriers" (2)

ISO 2709 specifies that a bibliographic record on tape should consist of a fixed length record label containing general information on the record, a datafield directory specifying the position of each data element within the record, and any number of variable length datafields, separated by a field separator (i.e. a special character) and terminated by a record separator.

As indicated in its scope, ISO 2709 specifies nothing more about data elements than that there may be any number of them and that they may be of any lenth (although the total length of a record is usually required to be no more than 2048 characters). This standard is used in most bibliographic data exchange

applications, except by a number of large commercial data-base systems for organizational, historic and economic reasons. ISO/TC46 is now working on a revision of the 2709 Standard. This is proving to be a difficult and time consuming task, due to differing views on implementation and processing in general and, of course, due to the economic aspects involved with the large investments in software for existing implementations.

ISO 2709 is not what is generally regarded as a bibliographic exchange format. As mentioned, it is no more than a logical structure specifying the lay-out of a record on the physical exchange medium. A bibliographic format in the usual sense is an implementation of the logical structure. Such an implementation has to specify:

- the elements, i.e. the fields and subfields to be distinguished in the bibliographic description (in theoretical terms: the implementation format specifies the set of domains from which the descriptive attributes are to be drawn)
- the identifiers for these elements, i.e. the codes which instruct the computer as to the type of element involved
- structural aspects, e.g. sequential or hierarchic relationships between elements
- processing information, e.g. indicators specifying repetition of data elements, number of non-filing characters, etc.

One of the most important things to understand about formats is that they do not only specify what information you put into the record, but also what information you can get out of it again. Here we come to one problem of format compatibility. If two formats use different identifiers for the same element, a mapping has to be established between the two (or two mappings between both exchange formats and the internal format of the processing system). This is obviously less efficient than having the same identifier in both formats, but it can be done. But if one format identifies an element and the other does not, a mapping cannot be established and the information is lost. Not identifying an element does not mean that the information provided by that element is not in the record. Imagine format A specifying title + subtitle as a whole, and format B identifying title proper and subtitle as two distinct elements (e.g. as subfields within the title field). Translating format B into format A is done by joining title and subtitle into one element. Translating format A into format B just cannot be done, as it is impossible to get title and/or subtitle out of format A separately.

Format compatibility is an economic necessity for efficient processing and optimal data exchange. One seemingly obvious solution, i.e. an extremely detailed general format, is not the optimal solution because it presents other economic problems, especially at the input side (3). Different users require different levels of complexity and definition in their formatted bibliographic records. A truly general exchange format will have to allow for various levels in a uniform way. Incompatibility between different levels can only be solved by standardizing users. I do not believe this would be a justifiable procedure.

Another major aspect of bibliographic exchange formats is that they are always implicitly or explicitly based on a set of cataloguing rules (or rather: description rules). Such rules do not only specify the various descriptive elements, but they also define them. Very often, different sets of cataloguing rules specify the same elements, but define them differently. For instance, the element 'title' may be defined as title proper in one format and as title proper + subtitle in another. Moreover, there might not be agreement as to what information within a document is to be regarded as 'title proper'. Another example: one set of rules might allow for more than three authors, while a different set might regard a publication by more than three authors as an anonymous document. There are other categorization problems as well. Most cataloguing rules identify authors, editors etc. But there exist rules which identify primary and secondary intellectual authorship. There is no foolproof way of mapping one into the other.

I believe therefore that standardization of cataloguing rules is the most fundamental problem behind compatibility of bibliographic data exchange formats. The library community has gone a long way in solving this problem, as I shall point out later in this paper. These efforts are reflected in the large degree of format compatibility within the library field. The field of Abstracting and Indexing services is much less uniform in this respect. Compatibility between both fields is virtually non-existing. This leads us to another problem in the field of bibliograpic information exchange. For there are various different types of bibliographic records. Libraries produce catalog type records. Abstracting and Indexing services produce reference type records. Catalog type records can be further differentiated according to various document types (e.g. maps, serials, patents, music etc.). Reference type records can be at the analytical, monographic and collective level. As indicated above, bibliographic records are a function of description rules as applied to documents. Different types of description rules, often necessitated by differences between document types, produce different types of records and this explains to a large extent the variety of formats now in existence. Different organizational contexts (e.g. libraries versus A&I services) also contribute towards variety and hence incompatibility. Integrating different types of records into one general exchange format in order to achieve efficient exchange of information is proving to be one of the most difficult tasks in the field of format standardization. This task can only be accomplished through efforts at all levels, i.e. description rules, formats,and the organization of information exchange as a whole.

THE PRESENT SITUATION IN BIBLIOGRAPHIC EXCHANGE FORMATS

The first format I should like to mention here is the MARC format, or rather the MARC family of formats. MARC stands for MAchine Readable Cataloguing and was initially developed by the Library of Congress (4). MARC-LC has a firm background in library cataloguing practice. It is based on the Anglo-American Cataloguing Rules (AACR) and is a more or less faithfull translation of these description rules for monographs into a machine-readable structure. The fact that MARC-LC has given birth to quite a family of highly related formats demonstrates the underlying problems of format compatibility and standardization. Most members of the MARC family are national or regional alternatives. They have been developed by Great Britain, Canada, France (MONOCLE), Italy, Belgium, Denmark, Austria and a group of French speaking West European countries (INTERMARC) (5).

These formats differ from one another in mostly minor respects due to a number of reasons:

- different cataloguing rules and conventions
- different views on processing
- local and organizational aspects (e.g. the bilingual situation in Canada)
- more recent development, offering more sophisticated techniques (e.g. the linking indicator in INTERMARC)

In general, however, there is a fair degree of compatibility between these MARC formats. They all identify more or less the same elements and they use the same identifiers for most identical elements. A good example of format development based on compatibility is the INTERMARC format, which retains MARC-LC as a valid subset. This means that, apart from having to apply minor software changes in processing, MARC-LC records are compatible with INTERMARC, though not vice versa without loss of information.

Another type of offspring from MARC-LC are the MARC formats for other types of documents than monographs (e.g. serials, maps, etc.).

Within the library community there are, unfortunately, several other formats which are not at all MARC compatible. The most notorious example is the German format, MAB-1 (6), which has a completely different structure. MAB-1 is based on the German cataloguing rules (RAK) which are nothing like the Anglo-American Cataloguing Rules. Also, MAB-1 is not entirely compatible with the present version of ISO 2709 (7)(8).

To summarize, the situation within the library community is as follows. Compatible exchange of bibliographic data is usually 100% within any one country, i.e. there is usually at least a national standard. There is a fair amount of international compatibility within the group of MARC-type format users, especially betwee the United States, the United Kingdom and a number of smaller countries (e.g. the Netherlands accept MARC tapes in their national system and will be MARC compatible when producing their own tapes). This is essentially true for monographs. There is much less compatibility between bibliographic records for different document types, for which you need different formats. This is mainly due to the fact that libraries have a much larger organizational differentiation between document types than the Abstracting and Indexing services. However, MARC users are now developing a general format called UNIMARC which is universal in the sense that it can handle a large number of different document types (9). Countries such as Germany, with its highly individual format, fall more or less outside the mainstream of international bibliographic data exchange. But on the whole, libraries all over the world are now making good use of the principal of Universal Bibliographic Control, recognizing that exchange of bibliographic data makes much cataloguing either unnecessary or at least a great deal more efficient.

The situation within the field of Abstracting and Indexing services is rather different. A number of services fall entirely outside the scope of format standardization by not even using the 2709 tape format, e.g. Excerpta Medica and Chemical Abstracts Services (10). Most formats within this field have been developed for specific types of A&I services, e.g.

- INIS for atomic energy information (also used by AGRIS and ASFIS) (11)
- EUDISED for educational information (12) (28)
- IRRD for road research information (13)(14)

In addition, there are various national or regional exchange formats for documentary information, e.g. MADOK (W. Germany)(15)(16) and MEKOF-2 (member states of the Council for Mutual Economic Assistance - CMEA) (17)(18)(19).

So the situation within the Abstracting and Indexing field is one of diversity. There is a large amount of international data exchange, but very little interdisciplinary compatibility. In other words, inter-system connection is impeded by the lack of a general exchange format, and perhaps also by the reluctance to accept one. The underlying problems in this field are slightly different from those in the library field. There are no really uniform description rules within the A&I community. However, this is a much less problematic issue here than in the library field where cataloguing rules are about the most important thing to argue about. Moreover, an international standard for bibliographic references has recently been developed by ISO (ISO 690), and though it is not specifically meant for A&I publications, it might help to provide more basic uniformity in this field.

Differentiation between document types is virtually impossible in the A&I field, at least at the organizational level. This means that you need one format which is universal to all document types and bibliographic levels. This is an advantage and a problem as well. An extremely elegant solution has been found in the UNISIST Reference Manual (20), which is based on a document type / bibliographic level matrix structure. The Reference Manual is a general exchange format developed by UNISIST in co-operation with the Abstracting Board of the International Council of Scientific Unions (ICSU/AB) for use by the Abstracting and Indexing services (21). Unfortunately, it has not been universally accepted, although there do exist various implementations (22).

It has been necessary to differentiate between the library field and the Abstracting and Indexing field in this discussion of existing formats. I have already pointed out that there is very little compatibility between the two fields. In my opinion, this is a serious matter. We could provide extremely efficient service to information users by integrating documentary and library information systems. This would avoid the frustration of finding a relevant citation in a documentary data-base and then having to spend a lot of time, energy and wit in trying to locate it in a catalog, often in another department, with a different type of description, and all that with a fair chance of not being able to find it at all. From a technological viewpoint, integration of both types of systems is feasible. The high degree of incompatibility between both types of information surely is one of the reasons why such integrated systems still hardly exist. We shall have to look at future developments in order to obtain some hope for a type of service we really should be offering our users now.

NEW DEVELOPMENTS IN FORMAT COMPATIBILITY

As I already have indicated, universal exchange of bibliographic information requires standardization of the information itself and standardization of the structure and content designation of bibliographic records, i.e. of bibliographic exchange formats. A great deal of this work has and still is being done by TC46 of the International Standards Organization. Such standards as for biblographic characger sets, standard book numbers, abbreviations of bibliographic terms, translitteration schemes, country and language codes, etc. all contribute towards more efficient bibliographic data exchange. Another organization working in this field is the International Federation of Library Associations (IFLA) which has developed the International Standard Bibliographic Descriptions (ISBD). These specify the lay-out and interpunction of essential elements for bibliographic descriptions. There are several ISBD's for various types of documents. IFLA has also developed a general ISBD(G), which integrates a large amount of different document types (23). It has now been proposed that ISBD(G) be adopted as an official international standard by ISO.

An interesting and extremely important development is now being undertaken by UNISIST, the Information Program of UNESCO, following a recommendation made by the International Symposium on Bibliographic Exchange Formats held at Taormina, Sicily, in 1978 (24). This major project, in close co-operation with other organizations from the library and documentation fields such as ISO, ICSU/AB and UNIBID, is designed to produce a common bibliographic exchange format which can be implemented at various levels of completeness by various segments of the information community. The format will be based on a so-called Data Element Dictionary, derived from a large scale comparative analysis of all major existing formats. One of the most important aspects of this format will be that it will provide a type of flexibility which I proposed in a paper (3) a number of years ago, i.e. the allowance of various levels of standardization, together with a co-ordinating instrument, viz. the Data Element Dictionary which will be governed by ISO. This will ensure a means of effective compatibility without serving as a uniform straight-jacket for all concerned. Especially, it will encourage information exchange from and to smaller scale systems, which have found the existing formats too complex and expensive for their needs. Hopefully, it will also provide the flexibility needed to adapt bibliographic systems to various types and levels of user needs. For, after all, it is the type of information needed by the user that should determine the contents of a bibliographic record (25).

So where will the future be? A new format will have its own problems. It will have to be accepted and it will have to prove itself. In the meantime work on bibliographic data exchange still goes on. I myself am now involved in a computer experiment with vaguely structured information, i.e. bibliographic information in which elements are not explicitly identified. This invloves having to instruct the computer to interpret bibliographic information along more or less the same lines as human beings do. Perhaps in future the computer will do more things than we can expect at this moment. Other means of establishing links between information services have been proposed, e.g. interfacing mechanisms based on authority control (26). This would be one way of bridging the gap between library (catalog-) systems and A&I (reference-) systems. An example of more specialized developments is the bilateral format for exchange of bibliographic data between the US and the USSR, drafted by the US-USSR Research Group on Development and Testing of a Common Communications Format for Bibliographic Data Exchange on Magnetic Tapes, within the framework of the US-USSR Working Group on Co-operation in the Field of Scientific and Technological Information (27).

In any case, world-wide exchange of bibliographic information is already well on its way. The future of bibliographic exchange formats will probably depend a lot on technological developments. For instance, we may expect exchange on tape to give way to on-line data exchange through communications networks such as EURONET. I believe this will have major consequences for future format design.

REFERENCES

(1) International Standards Organization. ISO 2709 Documentation - Format for bibliographic information interchange on magnetic tape. Geneva : ISO, 1973

(2) International Standards Organization. Draft Proposal ISO 2709-Rev. 1, Documentation - Format for bibliographic information interchange on magnetic tape. ISO/TC46/SC4 doc. N 124 (1978-09-15).

(3) Mackenzie Owen, J.S. Format incompatibility and the exchange of bibliographic information : a comparative study. Paris : UNESCO, 1976.

(4) Avram, H.D., J.F. Knapp & L.J. Rather. The MARC II Format : a communications format for bibliographic data. Washington, D.C. : Library of Congress, 1978.

(5) The interchange of bibliographic information in machine readable form : papers given at the European Seminar held at Banbury, 12-16 May, 1974 / edited by R.E. Coward and M. Yelland. London : The Library Association, 1975 (Library Association Research Publication ; 17).

(6) Deutsche Forschungsgemeinschaft, Bibliotheksausschuss. Maschinelles Austauschformat für Bibliotheken (MAB I). Arbeitsstelle für Bibliothekstechnik, 1973.

(7) Kohl, E. Current developments in Germany and Austria. In: The interchange of bibliographic information inmachine readable form. London, 1975, pp. 24-32.

(8) Kohl, E. Areas of possible incompatibilities in the international exchange of bibliographic information in machine readable form based on a compatibility study for MAB 1 data. In: Towards a common bibliographic exchange format? Budapest, London : 1978, pp. 52-58.

(9) International Federation of Library Associations (IFLA). UNIMARC : universal MARC format. London : IFLA International Office for UBS, 1977.

(10) Rowlett, R.J., R.L. Eigington & J.L. Wood. Standard distribution format : Chemical Abstracts Service. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 105-111.

(11) Dragulev, V., E. Ruckenbauer & D. Marek. The INIS format : an existing operational exchange format. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 122-131.

(12) Linford, J. The EUDISED format : a structure which will contain any type of formalized document record. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 115-118.

(13) Organisation for Economic Cooperation and Development (OECD). Specification for magnetic tape transfer of bibliographic information in the IRRD scheme. 3rd edition. Paris, OECD, 1977.

(14) Mongar, P. The International Road Research Documentation System's experience and views in bibliographic exchange formats. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 112-114.

(15) Zentralstelle für maschinelle Dokumentation (ZMD). MADOK : Magnetband-Austauschformat für Dokumentationszwecke. Munich, Verlag Dokumentation, 1977.

(16) Hitzeroth, Ch. MADOK - the German magnetic tape exchange format in the field of documentation. In: Towards a common bibliographic exchange format? Budapest, London, 1977.

(17) Mezdunarodnyj kommunikativnyj format - MEKOF 2. Soderzanie i sposob predstavlenija dannyh v zapisi dlja obmena informaciej na magninoj lente (NTP MCNTI 19-77): Specifikacija elementov dannyh. Moskva, MCNTI, 1977.

(18) Skripkin, V. & V. Kodola. The communication format in the international system of scientific and technical information of the CMEA countries. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 162-171.

(19) Vajda, E. Some ideas about the compatibility problems of bibliographic data interchange formats, based on a comparative anlysis of the MEKOF-2 format. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 172-192.

(20) Reference manual for machine-readable bibliographic descriptions / prepared by the UNISIST/ICSU-AB Working Group on Bibliographic descriptions with the assistance of ICSU and ICSU-AB member services ; compiled by M.D. Martin. Paris, UNESCO, 1974.

(21) Dierickx, H. The UNISIST Reference Manual for Machine-Readable Bibliographic Descriptions : current and anticipated developments. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 147-161.

(22) Hopkinson, A. The use made of the UNISIST Reference Manual : a brief history. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 49-51.

(23) International Federation of Libraray Associations (IFLA). ISBD(G) : General International Standard Bibliographic Description : annotated text. London, IFLA International Office for UBC, 1977.

(24) Towards a common bibliographic exchange format? : International Symposium on Bibliographic Exchange Formats, Taormina, Sicily, 27-29 April, 1978, organized by the Unisist International Centre for Bibliographic Descriptions (UNIBID) in cooperation with ICSU-AB, IFLA and ISO and sposored by UNESCO : Proceedings / edited by H. Dierickx and A. Hopkinson. Budapest : OMKDK-TECHNOINFORM, London : UNIBID, 1978.

(25) Mackenzie Owen, J.S. The fundamentals behind information exchange : an approach to optimal solutions. In: Towards a common exchange format? Budapest, London, 1978, pp. 9-12.

(26) Malinconico, S.M. The co-ordination of bibliographic control. In: Towards a common bibliographic exchange format? Budapest, London, 1978, pp. 61-67.

(27) Gove, N.B., T.E. Hughes, G.D. Matiushin & N.V. Turtanov. A common communication format for exchange of bibliographic data between the USA and USSR. In: Towards a common bibliographic exchange format? Budapest, London, 1978.

(28) Council of Europe. EUDISED : standards, formats, character representation. Strasbourg, Documentation Centre for Education in Europe, 1973.

KINDS OF ACCESS TO UNCLASSIFIED LITERATURE

by

C P Auger
Lucas Group Services Limited
Lucas Research Centre
Shirley Solihull
West Midlands B90 4JJ

SUMMARY

The paper reviews the nature of unclassified aerospace literature, and examines what categories are available and how they are organised. The preponderance of reports literature is noted, and current awareness and on-line services receive particular attention. Consideration is given to the great variety of users seeking access for one reason or another, and the different routes open to them, from general institutions such as public libraries to specialist information centres such as that at the Royal Aircraft Establishment. Finally attention is turned to some of the problems which need to be overcome if access is to be improved.

DEFINITIONS

The term classification has several meanings, but here refers to the classifying of documentary information into groupings according to the stringency of the measures to be taken to prevent its falling into the hands of an enemy or potential enemy. Typical designations are Top Secret, Secret, and Confidential. By inference therefore, and for the purposes of this paper, Unclassified means, in the context of documents, not subject to a security classification.

The other definition which needs stating at this point is that for aerospace itself, and the term may be taken to mean an industry whose products world-wide include civil and military aircraft, helicopters, aero-engines, guided weapons, hovercraft, space vehicles, and a whole range of aircraft and airfield systems.

WHAT SORT OF LITERATURE IS AVAILABLE?

Unclassified aerospace literature can be divided into a number of categories depending on the depth of understanding and degree of specialised knowledge necessary to make the fullest use of it. AGARD for example is concerned primarily with the exchange of scientific and technical information; other bodies are more interested in commercial and marketing information; others again in the historical, educational and strategic aspects; and of course aerospace literature is of considerable appeal to the general reader, whether as a schoolboy enthusiast or a regular airline traveller.

The wealth of aerospace literature and the great interest in it reflects the often mentioned point that a progressive industrial society needs a pioneering spearhead technology to stimulate growth, promote innovation and set standards of excellence for a country's entire range of manufacturing, production and trading activities. Factors which contribute directly to the volume of aerospace literature produced and disseminated, especially reports literature, are the immenseness of the projects involved; the employment of large numbers of highly qualified, highly skilled personnel; the importance of aerospace in terms of national security; and not least the ruling that contractors working on government projects are required to report on progress at regular intervals. To this list can be added public concern over safety, noise and pollution.

Secrecy has always been a major feature of aerospace information, and many documents stay inaccessible to those who cannot demonstrate a need to know. Nevertheless there still remains a large body of aerospace publications which is actively promoted to benefit the aerospace industry in general by making available a common core of knowledge, and secondly to try and justify and recoup some of the costs involved by promoting product spin-off and technology transfer.

HOW IS THE LITERATURE ORGANISED?

The forms of publication in which aerospace information appears are no different from those of any other major branch of knowledge, namely books, journals, translations, trade literature, standards and specifications, technical manuals, patents, and reports. What is different is the concerted effort which has been made to ensure such publications are readily accessible, and as a consequence the abstracting and indexing services associated with aerospace literature can be held up as a model to other areas of science and technology.

Another major difference is the preponderance of reports literature. Indeed it can be argued that the history of reports literature coincides almost entirely with the development of aeronautics and the aircraft industry. The series of reports usually accorded the honour of being first in Great Britain is the R&M (Reports & Memoranda)

series of the Advisory Committee for Aeronautics, now the Aeronautical Research Council, which began appearing in 1909. In the United States the aircraft industry has been represented continuously by the National Advisory Committee for Aeronautics (NACA), now of course the National Aeronautics and Space Administration (NASA), which issued its first report, On the behaviour of aeroplanes in gusts, in 1915.

Reports have since spread to other areas, notably nuclear energy, but still form a large and characteristic part of the aerospace literature.

As befits the country with the largest and most comprehensive aerospace industry in the Western world, the United States had led the way with two major abstracting publications. The first of these is Scientific and Technical Aerospace Reports (STAR) which is published twice a month as a guide to current technical reports issued by organisations around the world. STAR began publication in 1962, and since then some 325,000 citations have been listed. New citations are being published at a rate of 24,000 per year. In detail, STAR announces technical reports issued by NASA and its contractors, by other US government agencies and their contractors, and by both domestic and foreign companies, universities, and research organisations.

STAR reflects the interests of the NASA Scientific and Technical Information Branch (STIB) and its coverage embraces the basic and applied sciences relating to aeronautics and space research.

Despite the careful specification of NASA's subject interests by means of annotated scope notes, it has often been remarked that STAR is full of surprises. Thus for example it is possible to find details on establishing habitability factors for the design of office environments (N79-10744).

STIB is also concerned with the announcement of published documents other than reports, and accordingly sponsors the publication of International Aerospace Abstracts (IAA) by the American Institute of Aeronautics and Astronautics (AIAA).

IAA lists all journal articles, conference papers, books, and other forms of published literature, and the object is to complement rather than duplicate STAR, a task which with rare exceptions it manages to accomplish. IAA began publication in 1961 and by 1977 had presented around 500,000 citations of aerospace literature. Currently about 35,000 citations are announced each year.

Both STAR and IAA have been developed into a combined on-line data base which is available in Europe, but only to accredited users who enter into a tripartite agreement with the European Space Agency (ESA) and NASA, via the ESA-RECON installation at Frascati.

The NASA file is unique in that no royalties are levied - instead NASA requests users of its file to contribute on a regular basis suitable documents for inclusion.

In passing, it should be mentioned that the term RECON refers to remote console. In the United States, the data bases are available through NASA-RECON, Washington.

WHO WANTS ACCESS?

The range of users seeking access to unclassified aerospace information is as diverse as the forms of literature itself. The categories include research and development workers, performance engineers, company executives, university lecturers, students, and members of the general public. Each person will have his own reasons for wishing to consult the literature, for example the advancement of an experimental project, the establishment of design criteria, the assessment of a market, the submission of a thesis, or the study of a particular type of aircraft.

Detailed studies of specific categories of users have been presented at an earlier AGARD meeting (1), and include the contributions of J R Sutton (2) on the information requirements of engineering designers; Margaret O Sheppard (3) on users at the Aeronautical Research Laboratories, Australia; and Harold E Pryor (4) on methods used by NASA to evaluate user satisfaction.

HOW ACCESS IS ACHIEVED

Ease of access will depend to a large extent on the user's background. People working in research and development establishments and in large industrial companies will have the problems of identifying and obtaining literature considerably eased through the services of technical libraries and information departments. Lecturers and students in universities and colleges will be able to make use of main library and departmental collections, whilst the general reader's first resort will be the local public libraries, in particular the larger reference collections.

By using the appropriate channels, access can proceed from the local or regional level to national institutions, in the case of the United Kingdom the British Library Lending Division (BLLD) and the Science Reference Library (SRL).

The practical business of gaining access to aerospace information can be broken down into two components - identifying the publications required for the purpose in hand, and acquiring copies on loan or for retention.

Identification may be accomplished by consulting library and information specialists, by studying guides to the literature, and by making use of current awareness services such as STAR and IAA mentioned above. The widespread introduction of on-line searching facilities has considerably reduced the period taken to process enquiries, whilst at the same time it has greatly increased the comprehensiveness of the searches which can be undertaken.

In the particular case of aerospace literature, access to the ESA-RECON installation in the United Kingdom can be arranged through DIALTECH, the technical information service provided by the Department of Industry's Technology Reports Centre, (TRC) Orpington. Users in other ESA member states have similar access through the national centres nominated by ESRIN.

Details of experience with ESA can be found in many recent papers, including S Olmer - Court (5) on the French connection; L Maat (6) on the system at the Technical University, Delft; R Gülich (7) on experience at Sulzer AG; and S K Kumar and V A Kamath (8) on access to data bases in India.

Experience has shown that it is essential to impose an intermediary between the literature user and the on-line service. Usually the intermediary is an information specialist who can on the one hand exploit the interactive nature of the facility and make full use of the great number of access points, and on the other hand overcome the disadvantages of systems difficulties and data base inconsistencies. A further reason for using a trained intermediary is that an untrained user may inadvertently run up a large bill in his attempts to find references which match the requirements of a badly framed question.

Once references have been identified as relevant, the question of document provision crops up. References to published literature usually mean originals can be obtained on loan through whatever library channels are more convenient, or purchased through the book trade or official agencies. Because data bases provide full and detailed bibliographic information, problems of verification and checking are not normally encountered.

Items identified as reports however can present difficulties which are best resolved by applying to institutions which are known to specialise in such documents and to possess large reports collections. The TRC is one such source, another BLLD and a third SRL. Confirmation of holdings in the case of BLLD can easily be obtained for major reports series by consulting the latest edition of its publication "Current Serials Received", wherein are quoted the shelf marks for AD Reports, NASA publications of all types, PB Reports, and many more.

Again similar collections are available in ESA member states, although many European organisations prefer to apply direct to BLLD for a wide range of literature.

In the United Kingdom, a key collection of reports devoted specifically to aerospace is that maintained by the Royal Aircraft Establishment (RAE), Farnborough. Firstly it contains the research work of RAE itself, as recorded in Technical Reports and Technical Memoranda. In addition the research reports of other organisations all over the world are of great interest and use to RAE staff, and so are collected as well.

RAE make the point that the procedure for authorising the release of its own reports outside the Ministry of Defence (of which RAE forms part) is necessarily complex, even though the documents in question may be unclassified. The arbiter in such situations is the Defence Research Information Centre (DRIC), neighbour at Orpington of the TRC. RAE will advise on the availability of other reports in its collection.

PROBLEMS AND REMEDIES

A great deal of effort has gone into making access to unclassified aerospace information as easy and as efficient as possible. Providing the user for his part takes the trouble to think out and present questions in as precise a manner as possible, the system is available to provide speedy and relevant answers.

Three major areas of difficulty face the providers of information. Firstly, the need to educate the user so that he presents his requests logically and concisely, providing for example details of sources already known or tried, limitations on the period of interest, back-ground information likely to be helpful, and terminology characteristic of the subject under review. Only when the user knows clearly in his own mind what he wants and can be persuaded to take the information specialist fully into his confidence, will the best results be achieved. Admittedly many seekers after information are genuinely not able to refine their ideas since they are not quite certain what it is they are looking for until they actually see it. Nevertheless they can be encouraged to try.

Secondly people seeking access to information need to be aware of the importance of accurate data. Documents misquoted and authors' names misspelled are but two of the hazards which can prevent questions reaching successful conclusions.

Finally the user needs to realise that his and his alone is the task of understanding and digesting the information he has succeeded in gaining access to. Many readers are daunted by the sheer volume of the material they have unearthed and far from appreciating the efficient and rapid systems which have made it possible, are frequently moved to comment that they wished they hadn't asked!

APPENDIX

One of the first steps in the preparation of this paper was to conduct a search of the appropriate RECON data bases to see whether the subject of classified aerospace information had been examined, by other reviewers. It was felt that classified information ought to be easier to define than unclassified material, and that any discussion of such documents would inevitably provide an indication of the freedom of access to material not subject to security restrictions.

Accordingly a search was run in the NASA data base covering the period 1961 - 1979, using the following strategy : (1) REPORTS or (2) LITERATURE or (3) DOCUMENTATION or (4) INFORMATION RETRIEVAL, making set (5), which was combined with (6) CLASSIFYING and (7) SECURITY. No hits were registered with (5) and (6) and (7). Combining (5) and (6) only gave one reference, a report on the application of LANDSAT imagery. When (5) and (7) were combined, a total of 33 hits was indicated, but the papers themselves proved unhelpful, and the first couple of titles were concerned with the security of the national energy plan and a signature scheme for computer security.

Attention was then turned to the NTIS data base over the years 1969 - 1979, using the terms (1) DOCUMENT or (2) REPORT or (3) INFORMATION RETRIEVAL to make set (4), which was then combined with (5) CLASSIF and (6) SECURITY to give set (7), with an indicated total of seven hits. However when an extra term (8) AEROSPACE was added, no hits were indicated. Thus it was decided to see what set (7) contained, and the documents retrieved covered the following topics : solid waste composition and emission factors; military message experiments; the delivery of social services to elderly persons; Multics security enhancements (the Honeywell Multiplexed Information and Computing Service); a security compliance study of the Air Force Data Service Center; papers from the Classification Management Journal; and the Special Work Project for the Unemployed.

The most relevant of the above items was the collection of papers from the Classification Management Journal (AD-A025 341), the controlled and uncontrolled indexing terms for which at least confirmed that the strategy used was the correct one. However the search as formulated did not achieve the desired objective, namely the identification of papers on classified aerospace literature. No doubt other search strategies might have produced a different result.

REFERENCES

(1) The problem of optimization of user benefits in scientific and technological information transfer. NATO Advisory Group for Aerospace Research and Development. Neuilly sur Seine, AGARD 1976, 126p illus. tables. (AGARD Conference Proceedings No 179).

(2) J R Sutton. Information requirements of engineering designers. (AGARD Conference Proceedings No 179, 12-1 - 12-8).

(3) M D Sheppard. User response to the SDI service developed at Aeronautical Research Laboratories, Australia (AGARD Conference Proceedings No 179, 10-1 - 10-9).

(4) H E Pryor Listening to the user : a case study. (AGARD Conference Proceedings No 179, 11-1 - 11-7).

(5) S Olmer-Court ANRT, the national connection to the European Space Agency. (Information et Documentation, (3) Sep 76, 29-34).

(6) L Maat. The ESA/SDS information system at the Technical University, Delft, (Open 8(6) Jun 76, 304-308).

(7) R Gülich. Experiences with on-line searching (Nachr./Nouv./Notiz. 53(3) 1977, 97-107).

(8) S K Kumar, V A Kamath Access to world data bases and ESRIN/RECON on-line demonstration in India. (Ann. Libr. Sci. Docum., 23(3) Sep 76, 205-212).

Full Text Handling - a critical review

by Dr. R. P. L. Jones,
Computer Science Division,
Atomic Energy Research Establishment,
Harwell, Oxon OX11 ORA.
G.B.

ABSTRACT

This paper sets out to demonstrate the capabilities inherent in full text information retrieval or free text retrieval as the author prefers to call it. The elements of free text systems are defined and a description is given of the basic retrieval and display functions that will be common to most systems. The paper goes on to look at the circumstances allowing effective information retrieval using the free text approach. In particular comparisons are made with other mechanisms for information retrieval, notably the thesaurus based, keyword systems. The first part of the paper concludes with an examination of the strengths and weaknesses that are likely to occur in free text packages. The problems of retaining currency and accuracy in rapidly changing information environments are described and retrieval performance on large data bases is also discussed.

The second part of the paper draws upon the author's experience with the *STATUS* information retrieval system, to examine the wider application of the free text approach. Extended features are described to demonstrate how broad the application possibilities can be. Some comparisons are drawn with more traditional data base management systems in this context.

The paper describes the capabilities of free text systems when used as a nucleus for integrated information systems. Such systems allow not only ad hoc enquiries but make possible the dissemination of information through numeric and text manipulation using report generation and computer typesetting facilities. The capture of source information is discussed and the importance of linking to word processing systems is described.

The paper closes by suggesting that free text information systems are being evolved to provide solutions to a wide range of informational problems, traditionally being tackled by different software techniques. Furthermore, such systems open up the possibility that information scientists can design their own information layout and manage it in a retrieval system environment that is tailored to their needs, without having to involve anyone in system analysis or software development.

THE NUMERIC AEROSPACE DATA: Problems of Evaluation, Handling and Dissemination

G. K. Hartmann
Max-Planck-Institut für Aeronomie
3411 Katlenburg-Lindau 3, P.O.Box 20
Germany

SUMMARY

The main (numeric) data sources QI in science are till now scientific journals, books, tables, and reports ("Data from Literature" - in contrast to data about literature: "Bibliographic Data"). The advent of the new digital measuring and data processing techniques produced a second source QII of fast growing importance (data from measurements, from model calculation by computers, and from statistical inquires - briefly: "Measured Data"). The large data growth rates due to the very recent technical progress give rise to a slow increase of a third source QIII ("Picture Data"). Since the evaluation-, dissemination-, and handling-problems of Aerospace data from QI are very small - and principally solved - compared to those considering QII and QIII they will be disregarded here. Since the problems with QIII are even more complex and more recent than those of QII, they have to be disregarded, too, due to the short time available here. The organic specialized information growth led not only to the beginning of the second data source QII but also to the necessity to redefine or subdivide the terms data acquisition and data processing. For example it became necessary to introduce the new terms "raw data acquisition" and "data preediting" as well as to subdivide the term data processing, in context with data evaluation procedures. In context with dissemination and handling procedures we are now faced with the problems to distinguish between small, medium, and large information systems.Due to the fact that most of the numeric Aerospace Data ("Direct Aerospace Data" from source QII) are non reproducible time series data - often also space dependent - with tremenduous growth rates it will be extremely difficult and time consuming for any scientist ("secondary user") and information system (documentation personnel) not actively working in the specific field, to evaluate, disseminate and handle those data with a high degree of efficiency and reliability. Thus it is very likely that also in the future numeric Aerospace data exchange activities will by far dominate the pure handling activities despite any technical progress.

I. INTRODUCTION

The fact that we have to talk about such general problems in context with Aerospace data is due to the tremenduous data growth rates (information explosion) which we observe in the industrialized nations during the last two decades.

There are two principal information problems [3, 4, 5, 6, 7, 8]:

1. Information can be accumulated, time (e.g. human life time) cannot be accumulated.
2. Further information growth rates will increase the tension between structural differentiation (i.e. branching which is a consequence of so called organic growth) and integration (i.e. striving for better overall views, which are a need of our present societies and which are a consequence of new scientific results and new technical tools).

Branching now often leads to an organisational and institutional separation between Data Acquisition, Data Documentation, and Data Processing Activities as well as between Research and Service Activities. This often leads to various unexpected, negative consequences. Documentation activities [1, 2, 8] will play a key role in avoiding and/or reducing the negative separation effects. (Documentation means: the systematic collection of literature in a broader sense, the analysis of the facts and the storage of its principal elements with the intention of providing a fast and complete retrieval of the information upon request.)

The main (numeric) data sources QI in science are till now scientific journals, books, tables, and reports ("Data from Literature" - in contrast to data about literature: "Bibliographic Data"). The advent of the new digital measuring and data processing techniques produced a second source QII of fast growing importance (data from measurements, from model calculation by computers, and from statistical inquires - briefly: "Measured Data"). The large data growth rates due to the very recent technical progress give rise to a slow increase of a third source QIII ("Picture Data"). Since the evaluation-, dissemination-, and handling-problems of Aerospace data from QI are very small - and principally solved - compared to those considering QII and QIII they will be disregarded here. Since the problems with QIII are even more complex and more recent than those of QII, they have to be disregarded, too, due to the short time available here.

Talking about Aerospace data in general and numeric Aerospace data in special we have to distinguish in both cases between the following two categories:

A. Direct Aerospace data: Data that stem from the investigation of the Aerospace.

B. Indirect Aerospace data:
   a) Data about any device for the investigation of the Aerospace
   b) Data about any type of Aerospace transportation and/or communication/navigation system

Here we will only deal with numeric data from category A, respectively QII, and denote them for the

sake of brevity "numeric Aerospace data". Already now they represent the major portion of all the available Aerospace Data.

II. EVALUATION PROBLEMS OF AEROSPACE DATA

1. Problems due to different properties of data

Generally speaking, any information that can be processed by a computer can be regarded as "data". However this general meaning produces many misunderstandings in almost all practical cases when we are faced with more detailed data activities. Thus we have to specify this term [4, 7, 8].

Tables 1a and 1b show the main areas of data documentation and some data filing aspects.

More than 90 % of the numeric Aerospace data belong into the category of time and space dependent non reproducible data, which introduces special problems for any data evaluation and interpretation activity.

Fig. 1 which shows the elements of empirical science will clarify this aspect from another point of view. It becomes clear that all data that belong into the category "reproducible" belong here into the category "conditions controlled by man"; "non reproducible data" belong into the category "uncontrolled conditions". This presentation shows simultaneously that most of the interesting relations deduced from numeric Aerospace data have to be obtained by statistical methods. There are two principal statistical methods of treating large amounts of data a) Frequency distributions, b) Descriptive statistical parameters. An additional problem is introduced by the fact that very many of the just mentioned numeric Aerospace data are time series data. (A time series is a set of observations at different times. Each observation or measurements consists of a quantitative number that describes the measurement as well as of a second number that describes the time at which the measurement was carried out.) If we deal with time series (data), then there are till now no well developed and tested methods of inference (conclusion), applicable to large areas, like those which one knows from analytical statistics. This is mainly due to the following fact: Concerning time series data we find that adjacent observations are mostly correlated, e.g., they are not independent from each other. This so called "series correlation" does not allow the generally used estimations of standard deviations and troubles the "range-estimations" or check of hypotheses. This implies that the evaluation or interpretation of those (non reproducible) time series data requires a much larger and longer practical experience than the evaluation of reproducible data. This is very often not enough considered in context with the planning of those information systems separated from the relevant research institutes and dealing with (non reproducible) time series data. The problems increase further if we have to consider simultaneously stronger spatial variability of the data.

2. Data Acquisition and Data processing problems

Despite the fact that in general numerical data are the most objective information sources and therefore are the most important basis for rational decision finding processes, one must not forget that very likely the majority of decisions and actions are based on emotions, i.e. they depend on subjective information [4, 7, 8]. Together with the tremendous information growth rates, especially for QII and QIII, and the still increasing software-hardware - gap this leads to the increasing use of subjective information. This ultimately implies that decisions reached on a golf course or by drinking a beer with just a second person, for example, might dominate by far those based upon more objective data that stem from specific information systems. In the long run this will strongly affect all kinds of official information channels, especially the large Information Analysis Centers (IACs) [1, 2], the WDCs (World Data Centers) [10], as well as all larger data acquisition systems, e.g., SELDADS [9].

The most measured data from category QII are stored digitally these days and mostly without any explanatory remarks. This "minimum documentation" makes the data only usefull for an extremely small user-community, in general only the principal investigator (PI), who just matched the procedure to his own special needs. These days attempts are made that part of these data - at least those that stem from very expensive experiments or those that are of increasing interdisciplinary interest - is made available by additional documentation procedures for a broader user-community, e.g., with the help of various IACs. This would be a large step beyond the till now predominant data-exchange or cooperation schemes within one specific field of science. Already now there become visible quite a number of problems e.g., the unconscious and/or conscious data misuse by secondary users, data protection problems as well as many administrative and economic problems.

Due to the large data growth rates we have to subdivide not only the term Data Acquisition - see fig. 2 - but also the term Data Processing - see fig. 3. Fig. 2 shows also the meaning of the newly defined term Data Preediting and shows further a new term, defined by the author namely: Datography, which is used in analogy to the term Bibliography. (Datographies give that minimum of information - e.g. type of data, accuracy, calibration, sampling rates, formats etc. - which is required that "secondary users" can use numeric data - generated by principal investigators - with minimum chances of misinterpretations.)

In the context with all Data preediting activities Video-Graphic-Communication and Documentation Systems (VIGRODOS) will play a very important role [4, 7, 8].

Special data documentation steps or datographies have to be performed on each step from the data acquisition procedure to the final data processing (data evaluation and interpretation). The broader the "secondary" user-community will be and the less "background" information these user have, the more detailed the datographies have to be written. Very detailed datographies however lead to two problems on the one hand it is very time consuming to write those detailed descriptions and on the other hand it also very time consuming to study very detailed descriptions. The less detailed a datography the smaller

the number of additional users!! Thus we have to envisage a compromise between the maximum wanted number of "secondary" users and a necessary minimum text for the datographies. This will lead in the documentation area to a number of principal problems. If information-systems deal with large amounts of data and they want to supply various related services, then in addition to datographies they have to introduce another "activity", which is denoted as "Data-Catalogue" by several information handling agencies, e.g., by WDC-A in Boulder, Colorado, USA (fig. 3, Table 5).

3. Some examples for the production of large amounts of data

a) Space Transportation System (STS)
The STS-program plans about 500 flights between 1979 through 1991 for the Space Shuttle and the Spacelab. A maximum data flow of 5×10^7 bits per second is envisaged. This corresponds the "production" of a "DIN A4" file with 500 pages densly printed with text or tables. Thus for a 7 days flight period with 100 % duty cycle we could get 600 000 of such files with information from several scientific domains. After 12 years - 500 flights - this would be 30 000 000 of such files with 500 pages each, almost twice as much as the largest library of the US Congress (Library of Congress) stores today concerning the number of volumes.

b) Other programs
The just mentioned STS-program is only out of a larger number of data intensive scientific programs which are already in operation or which are under preparation, e.g. satellite-meteorology, Earth resource satellite programs, Aerospace programs etc.
In the 1976 report "Geophysical Data Centers: Impact of Data-Intensive Programs", the Geophysical Research Board of the U.S. National Academy of Sciences' National Research Council evaluated the impact of large scale geophysical programs on the U.S. and WDC-A. They found that the National Climatic Center had about 77.000 reels of digital magnetic tape. The National Space Science Data Center had about 41.000 reels, and the National Geophysical and Solar Terrestrial Data Center had about 600 tapes from only a few years of data acquisition of this type. These Centers, together with the National Oceanographic Data Center, had millions of feet of film records and cubic feet of paper documents. Further, a sampling of some 14 national and international data-collection programs then in progress indicated that these data loads would increase by some 10^{14} "bits" (equivalent to 2.5 million digital magnetic tapes). This is in addition to the continuing important Center roles of archiving data in photographic, graphical, and paper tabular forms - truly a formidable task [4].

c) Directly application oriented programs
For example the oil-companies need large amounts of geophysical data for their resource-technologies. Some of the very large companies have thus several millions of tapes with data. The progress of measuring techniques as well as the progress in evaluation techniques as well as the specific characteristics of time series data imply that these tapes will be processed not just once but several times spaced by several years. This leads to serious documentation problems, which have to be solved and which are not only due to the large amount of data tapes but also to their technical maintenance.
However there is nowadays in other fields than physics or geophysics a "production" of large amounts of data visible, due to the very recent developments in electronics etc. and due to very many model-calculations which are performed with the help of modern computers.

III. DISSEMINATION AND/OR HANDLING PROBLEMS

1. Data Exchange (exchange of measured data, source QII)

One of the most important activities in the Aerospace domain was and is the international cooperation of institutes, the performance of joint experiments and the exchange of data between institutes that performed the same or similar measurements. For a very large amount of these data international cooperation is one of the main postulates for an effective data interpretation. This mutual exchange dominated till now the whole data documentation scene in geophysics (Aerospace). This exchange took place directly between research institutes or via the WDCs etc. This was in contrast to most other areas in physics where the data, for example from laboratory physics, were available not only for the specialists but also for a wider user-community, that was often only interdisciplinary related to the subject. Just now there are some additional attempts initiated to try to make available parts of the geophysical data - at least those that stem from very expensive experiments or those that are of increasing interdisciplinary interest - to a broader user-community, e.g., with the help of newly established IACs. However, this produces not only economic and administrative as well as organizing problems but also a principal one as already mentioned. This implies also that one of the major tasks of the new information systems is "Advice" for secondary users who request non-reproducible data. This advice is getting the more difficult the more we deal with time series data.

2. Other problems

One major problem arises due to the fact that we mostly have to deal with various generations of data as is shown in table 2 presenting various steps of empirical, experimental sciences.

A further problem is due to the fact that large amounts of data can stem from principally different data sources and generations - table 3 - and can be used for different purposes, for example Research or Service Activities - table 4.

Another very important problem arises these days due to the fact that we have to deal with three types of information systems: small, medium, large - see table 5.

It seems important to mention that we have not only to distinguish between the various services that are supplied by the information-systems but also to keep in mind that we also generally deal with different

"generations" which lead to different "main tasks" for the systems.

In the first generation when a new field of research starts the main task of a system is the collection of data and to supply the information who is doing what and where. The data should in general be preedited but not compressed. We deal only with fairly small amounts of data. In the second generation when the field of science is well established we deal with a lot more information. Then the main task for the information-systems is the documentation. The data are mostly distributed in a slightly compressed form. In the third generation we deal with very large amounts of information (literature and data) and a wider distribution of "raw data" and/or preedited data by information centers is impossible. The data have to be either compressed or small portions have to be selected before they are mailed. Thus the compression of data and the compilation of data catalogues become now the main task besides the older already "classical" tasks.

IV. CONCLUSIONS

Due to the fact that most of the numeric Aerospace Data (Direct Aerospace Data from source QII) are non reproducible time series data - often also space dependent - with tremenduous growth rates it will be extremely difficult and time consuming for any scientist ("secondary user") and information system (documentation personnel) not actively working in the specific field, to evaluate, disseminate, and handle those data with a high degree of efficiency and reliability.

Thus it is very likely that also in the future numeric Aerospace data exchange activities will by far dominate the pure handling activities despite any technical progress.

V. REFERENCES

[1] Abbel, R., 1978
Dateninformation und -dokumentation -Einführung und Oberblick-
In: Deutscher Dokumentartag 1977 (03.10.-07.10.) in Saarbrücken, S. 133 ff.
Bearb.: Mathilde von der Laake, Peter Port (DGD-AIS), Verlag Dokumentation Saur KG, München, New York

[2] CODATA, 1977
CODATA (Committee On Data For Science and Technology)
CODATA Secretariat, 51 Boulevard de Montmorency, 75016 Paris, France

[3] Hartmann, G. K., 1977
Why do we need data, what principal data evaluation problems have to be considered?
Herausgeber: ZAED, Kernforschungszentrum, D-7514 Eggenstein-Leopoldshafen 2

[4] Hartmann, G. K., 1978
The information explosion and its consequences for data acquisition, documentation, and processing
World Data Center A. Solid Earth Geophysics, Report SE-11; Solar Terrestrial Physics Report UAF-65, U.S. Department of Commerce, NOAA EDS, Boulder, Colorado 80302, USA

[5] Hartmann, G. K., 1978
Betrachtungen über Energie und Information. ntz, Bd 31, Heft 8, S. 535-536

[6] Hartmann, G. K., 1978
Informationsexplosion - Grenzen, Folgen, Handlungsmöglichkeiten
Teil 1: Energie und Information. ntz, Bd 31, Heft 8, S. 568-571
Teil 2: Probleme des Informationswachstums. ntz, Bd 31, Heft 9, S. 654-657

[7] Hartmann, G. K., 1978
Numerical Data Growth Rate Problems
6th International CODATA Conference, Santa Flavia, Italy, May 22-25, 1978 (Proceedings in press)

[8] Hartmann, G.K., 1979
Documentation problems of large amounts of measured data
Edited by FIZ 4; Kernforschungszentrum, D-7514 Eggenstein-Leopoldshafen 2, Germany (In press)

[9] Williams, D., 1976
SELDADS: An operational real time solar terrestrial environment monitoring system
NOAA Techn. Report ERL-SEL 37, U.S. Department of Commerce, Boulder, Colorado 80302, USA

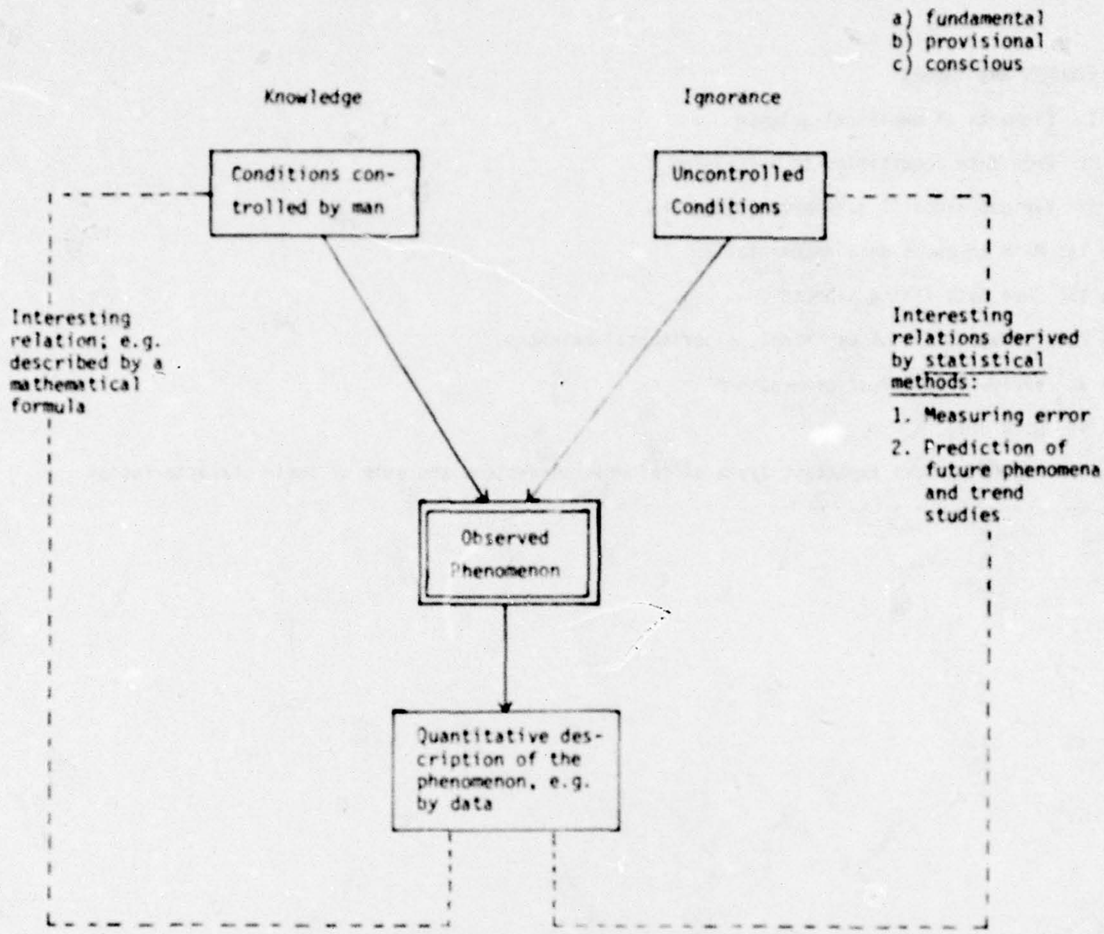
[10] NOAA EDS, 1978
Proceedings of the NGSDC Pilot Workshop on the information explosion and its consequences for data acquisition, documentation, and processing
NOAA EDS NGSDC, Boulder, Colorado 80302, USA

VI. ACKNOWLEDGEMENT

The author thanks the Max-Planck-Institut für Aeronomie, represented by the director, Prof. Dr. W. I. Axford, for the support. Many colleagues he thanks for critical discussions, especially at the WDC-A-STP at Boulder, Colorado, USA and at FIZ 4 at Karlsruhe, Germany.

VII. FIGURES AND TABLES

Fig. 1: Elements of empirical science



Remarks:

A. The ignorance about the uncontrolled conditions can be of three different types:
   a) fundamental type - if for instance the influence of a parameter cannot be determined in principle (time of nuclear splitting)
   b) provisional type - if we can hope that by the use of new information we can expand the range of controllable conditions
   c) conscious type - if we know of the existence of some uncontrolled conditions and disregard them, since regulating them is either of no interest or technically not feasible or too expensive.

B. There are two principal statistical methods of treating large amounts of data.
   1. Frequency distributions
   2. Descriptive statistical parameters

Fig. 2: From data acquisition to processing

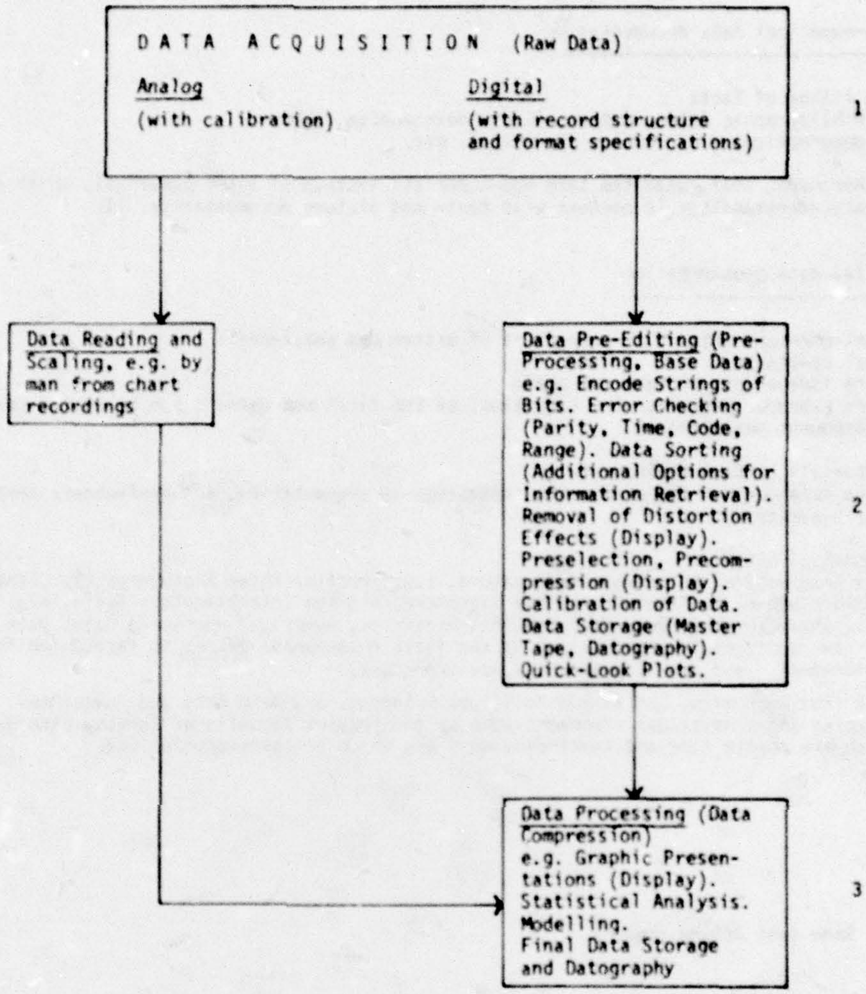


Fig. 3: Various steps of data processing

1. Data collection (only preedited data are collected)

2. Data collection documentation (compilation of datographies and preparation of data catalogues)

→ Archives and Data Bases

3. Data compression
   a. Graphical representation
   b. Statistical analysis
   c. Models

→ Publications (literature)

The third step, the data compression is again subdivided with an increasing degree of complexity.

Table 1a   Main areas of data documentation

I. Alpha-numerical data documentation

Systematic filing of facts
Examples: Bibliographic files, address files, merchandise lists, geographical distribution of insects, etc.

Remark: Hartmann, 1977, used the term non-numerical instead of alpha-numerical, which produced misunderstandings in context with text- and picture documentation. [3]

II. Numerical data documentation

1) Physical-chemical data, i.e. properties of matter and substances;
   Physical constants
   Data are independent of time and space
   Data are reproducible since the conditions of the first measurement can be reproduced for any subsequent measurements

2) Raw material- or Engineering data
   Data are dependent on the production technology of the material, its preliminary treatments and its "geometry"

3) Environmental data
   a) Data from natural matter and from nature, i.e. from the three Geospheres (1. Lithosphere, 2. Hydrosphere, 3. Atmosphere), the Biosphere, and the Interplanetary Space, e.g.: Geosciences, Biosciences, Medicine, Astrosciences, numerical Aerospace Data. Data are practically all non reproducible - conditions of the first measurement cannot be reproduced for any subsequent measurement - and are time and/or space dependent.
   b) Data from Economics, Social and Political Sciences, and from Arts and Humanities. Example: The statistical almanac issued by the Federal Republic of Germany give data which are mostly time and space dependent and which are non-reproducible.

Table 1b   Some data filing aspects

I.	Properties; characteristics e.g. time and/or space dependent	(1. Step)
II.	Origin; domain	(2. Step)
III.	Measuring methods	(3. Step)
IV.	Utilization, application	(4. Step)

Table 2 Various steps of empirical, experimental sciences

First generation

(Simple Data Technologies, SIDAT)

Data Acquisition (DA), Data Preediting (DP), Data Documentation (DD), Data Processing (DPr), Data Exchange (DE), and Data Handling (DH) are done at the same location and by one institution. Here we deal with a relativ small amount of data and mainly with pure research activities.

Second generation

(Intermediate Data Technologies, INDAT)

The specialisation and "branching" processes due to the organic information growth imply that the steps DA, DP, DD, DPr, DE, and DH, are getting complexer and are slowly institutionally and organizingly separated, i.e., they occur generally at several different locations, being accomplished by several different organisations. Here we deal in general with medium amounts of data. We still have more research than service activities. (Service means in this context the foundation of so called information systems for data handling.)

Third generation

(Recent Data Technologies, REDAT)

The above mentioned steps DA, DP, etc. are mostly completely separated, spatially, temporally, and organizingly. Here we deal with a large amount of data. Generally there are more service activities required than research activities. (Service for Research and Application.)
Foundation of Information Analysis Centers (IACs), Data-networks, and Data Base Management Systems (DBMS).
Here we reach at present the limits of the human information processing capabilities and capacities.

Fourth generation ????

There might later occur a leap across the just mentioned limits, possibly initiated by new ways of information experience and information understanding which amongst others will be supported by VIGRODOS-Systems (Video-Graphic-Communication and Documentation System; term defined by the author).

Table 3 Various "technical generations"

By using large amount of data one has to take into consideration various "generations" and sources.

A. Very many different, short-time sources

↓

Generally three technical generations have to be considered:

a) SIDAT (see Table 2)

b) INDAT (see Table 2)

c) REDAT (see Table 2)

↓

Are they comparable and how?

↓

Compilation of a data catalogue (datography etc.)

↓

Edition of an "electronic book" with the data (e.g. cassette for a mini-computer with graphic display)

↓

Multidimensional "still picture presentation e.g. on a television screen

B. Very few, long-time sources which produce time series data

↓

Generally three "time generations" (old, medium, and young) have to be considered. They are represented by:

a) Very coarse models

b) Improved models

c) Fine scale models

↓

Where lies the "data noise" threshold?

↓

Compilation of data for selected short time periods, with special datography

↓

Graphical representation of selected data (e.g. with cassette on a mini-computer with graphic display)

↓

Multidimensional "motion-pictures" (film, holograms, etc.) on television screen using also video mixing techniques

Table 4 Data Usage

Large amounts of data might be used for the following two purposes:

A. RESEARCH, (Science)

a) by the principal investigator (PI)

b) by the secondary user

↓

about 60 % of the data can be directly processed using computers and filter techniques according to "apriori knowledge" (redundant). Here we deal with the application of technical intelligence.

↓

about 40 % (rest) requires the direct interaction of human beings and their decisions, i.e. organic intelligence.
Only this part leads to real new knowledge.

B. SERVICE, (operational activities)

a) by the data producer

b) by the secondary user

↓

about 100 % of the data can be directly processed using computers due to "apriori knowledge"

↓

The data are used for prediction purposes and for improving the prediction schemes (the longer the time series the more we are moving from "frequency distributions" to "probability distribution). Only in this context we obtain really new information.

Table 5 The three most important types of information-systems and some of their characteristics

		Small information-systems	Medium information-systems	Large information-systems
I.	Main-Service	preedited, uncompressed data	slightly compressed and selected data	data-catalogues (preedited and compressed data only on special request. Store period > 10 years).
II.	Main-Technique	SIDAT + INDAT ●● (VIGRODOS)	INDAT + REDAT ●● (Data-Networks)	REDAT ●● (DBMS and Data-Networks)
III.	Main-User	Many individuals	Research groups	Large organizations
IV.	Number of users	Very large	Medium	Small
V.	Budget	Public and/or private	Public and/or industry	Public, industry, military
VI.	Store period	< 10 years	≈ 10 years	> 10 years ●
VII.	Amount of data	Small	Medium	Large
VIII.	Flexibility	Large	Medium	Small
IX.	System time constant	Small	Medium	Large
X.	Expenses	Small	Medium	Large
XI.	Computer	Small	Medium	Large
XII.	Accumulationproblem	Large	Large	Large
XIII.	Location	Decentralized	Decentr. and centralized	Centralized
XIV.	Ratio of formal to informal information	Small	Medium	Large
XV.	Ratio of research to service	Large	Medium	Small

●● see also table 2

● the uncompressed, preedited data are only stored longer (>) than 10 years, if they were requested during this period. Otherwise only the data-catalogues and the compressed data will be further stored.

The sign ≈ means "approximately", < means "smaller", > means "larger"

THE ROLE OF WORLD DATA CENTERS AND THE LUNAR AND PLANETARY INSTITUTE IN THE INTERNATIONAL EXCHANGE OF LUNAR AND PLANETARY DATA

Frances B. Waranius
Lunar and Planetary Institute
3303 NASA Road One
Houston, TX 77058

SUMMARY

The success of many lunar and planetary investigations has resulted in the accumulation of a mass of data in a myriad of formats and medias. The application of these data to comparative planetology, origin of the solar system, and potential industrial applications of space, has made it necessary for scientists from many disciplines to access the data.

Although much of this data is archived at the National Space Science Data Center (NSSDC) and made available to scientists world-wide through the World Data Center-A Rockets and Satellites (WDC-A/R&S), the collections of this data are so diverse that it is often difficult to select what is needed based on catalog information alone. In addition, there are other valuable data sources which are not at NSSDC.

The Lunar and Planetary Institute (LPI) has one essential goal...to bring the user and the data together. As a research support organization operated by the Universities Space Research Association (USRA), the Institute through its active Visiting Scientist Program, a balanced program of study workshops and topical conferences, and organized and supervised data collections, has assisted scientists, educators, and students to review, study, and obtain the data necessary to the pursuit of their research.

The concept of World Data Centers for the international exchange of data was initiated during the organization of the International Geophysical Year (IGY). Prompted by the problems which occurred in obtaining data from the observations of the Second International Polar Year (1932-33) the Comite Special pour l'Annee Geophysique Internationale (CSAGI), in September 1955, resolved that "observational data to be exchanged in accordance with the IGY programme shall be available to scientific institutions in all countries" and accordingly authorized "the establishment of at least three IGY World Data Centres, of which one will consist of a number of parts. Each Centre will be international in the sense that it will be at the service of all countries and scientific bodies." As a result of offers by the United States and Russia to establish centers for all IGY disciplines, it was decided to establish WDC-A in the United States, and WDC-B in the USSR; a third center, WDC-C is comprised of a number of discipline centers in various nations.

This new form of international cooperation--international exchange of data through World Data Centers organized for the IGY--was found to be very effective. Instead of having to address themselves to many national organizations, scientists could receive data necessary for scientific work directly from the WDCs. Thanks to the collection and exchange of data through WDCs, it became possible to investigate phenomena on a planetary scale and to study the interdisciplinary relationships among various phenomena.

The World Data Centers collect data and publications in the following programs:

- Solar-Terrestrial Physics
- Rockets and Satellites
- Meteorology
- Oceanography
- Glaciology
- Solid Earth Geophysics

World Data Center A (WDC-A) consists of the WDC-A Coordination Office within the National Academy of Sciences and seven subcenters at scientific institutions in various

parts of the United States. Five of the WDC-A subcenters are operated by the Environmental Data Service of the Department of Commerce's National Oceanic and Atmospheric Administration, one by the National Aeronautics and Space Administration (NASA), and one by the U.S. Naval Observatory. WDC-A follows the guidelines set down in the International Council of Scientific Unions (ICSU) Guide to International Data Exchange (1973). The fourth edition of this Guide is due the summer of 1979. The Guide is available from any of the WDC-A subcenters or the National Academy of Sciences. (1)

The Center with the prime responsibility for the exchange of lunar and planetary data is WDC-A/Rockets & Satellites (WDC-A/R&S) which is located at the Goddard Space Flight Center in Greenbelt, Maryland. This is contiguous to the National Space Science Data Center (NSSDC) which was established by NASA in 1965. NSSDC was commissioned with the responsibility to collect the scientific data obtained from space probes, satellites, sounding rockets, stratospheric balloons, and high altitude aircraft. It is responsible for the acquisition, organization, storage, retrieval, announcement, and dissemination of this information. The types of data archived in this subcenter include:

Reports of sounding rocket launchings
Reports of satellite and space probe launchings
Detailed descriptive information on spacecraft experiments
Scientific reports of limited distribution on results of experiments
Orbital elements and ephemerides which are of scientific interest and value

There are data from over 600 satellite experiments covering such disciplines as lunar and planetary photography; x-ray; gamma-ray, RF, UV, and visible astronomy; cosmic rays; magnetospheric energetic particles, plasma, and electromagnetic fields; interplanetary energetic particles; ionospheres; and meteorology and IR data. (2)

The number of experiments for which data were available at NSSDC grew from a total of 105 in 1967 to 225 in 1969. With the advent of the Apollo program, the total number of experiments climbed until in 1974 there were 548 experiments. Recent tabulation shows that this total rose to 793 in 1978. (3) Of this total, 116 experiments are from 18 planetary exploration spacecraft and 175 experiments are from 29 lunar exploration spacecraft.

Planetary data available to researchers includes more than 150,000 frames of photographic data, 600 reels of magnetic tape, 110 rolls of microfilm, and 1,400 sheets of microfiche. Lunar data available includes more than 152,000 frames of photographic data, 3,111 reels of magnetic tape, 1,500 rolls of microfilm, and 140 sheets of microfiche. (4)

Thus, even though the mechanism for the international exchange of space data did exist through the WDCs, it was recognized that this body of knowledge resulting from the massive technological and scientific effort known as the Apollo Program, would create special demands for interaction among the world's scientific community.

In a speech at the Manned Spacecraft Center (now the Johnson Space Center), on March 1, 1968, President Lyndon B. Johnson proclaimed:

> "As a further step toward joining hands with the world's scientific community, I want to announce that we will build facilities here in Houston to help the world's scientists work together more effectively on the problems of space. We are going to have a new Lunar Science Institute alongside this great Center. ... This new Lunar Science Institute will provide new means of communication and research for the world's scientific community. It will help unite the nations for the great challenge of space."

And so the Lunar Science Institute (LSI) began. It was formally established on October 1, 1968 by a NASA contract with the National Academy of Sciences (NAS). Renovation of a vacant old mansion which had been the home of a Texas millionaire in the 1930's was begun and on October 27, 1969, the LSI moved into its permanent quarters.

Since it was necessary to have an organization to permanently manage the LSI and possibly additional space-related facilities, the NAS convened two meetings of university representatives to discuss a new university consortium. Thus, the Universities Space Research Association (USRA) was incorporated in the District of Columbia on March 12, 1969; the first meeting of USRA institutional representatives was held in Washington, D.C. on June 9, 1969, and USRA assumed the management of LSI under contract to NASA on December 11, 1969. To accomplish the goals of international cooperation outlined by President Johnson, the LSI initiated several programs:

Visiting Scientist Program
Conferences, Symposia, Workshops
Publications
Library

The Visiting Scientist Program is an effective means of bringing together scientists from both domestic and foreign universities and industries who wish to begin or continue work on the results of the lunar and planetary missions. The Visiting Scientist Program provides for experimental and theoretical research and educational program development in the lunar and lunar-related sciences and in comparative planetology. Experimental research programs proposed by appropriately qualified individuals may be accommodated through utilization of laboratory facilities at the Johnson Space Center. Appointments vary in length from a week to as long as a year, with option for renewal at the discretion of the Institute's Director. Appointments range from the Associate Professor level through Post-Doctoral appointments, Graduate Fellows and Undergraduate Interns.

To promote the dissemination of information and research results, an active schedule of conferences, topical symposia, study workshops, and educational short courses, are conducted by the Institute's Symposia Office. The results of these meetings are edited and published by the Publication Office. Chief among these, of course, is the annual Lunar Science Conference (now the Lunar and Planetary Science Conference). This major conference which regularly brings together scientists from all over the world is held under the joint sponsorship of the Institute and the Johnson Space Center. The average attendance at this week-long meeting is usually about 650. Two major publications which represent the greatest single summary of the results of the lunar and planetary investigations are the abstracts of papers presented to the conference and the proceedings which each year have been a three-volume work averaging about 3,000 pages. The first of these conferences was held in 1970; the most recent, the Tenth was held in 1979.

Recent topical conferences have included a Workshop on the Thermodynamics and Kinetics of Dust Formation in the Space Medium, Conference on Origins of Planetary Magnetism, Workshop on Ancient Crusts of the Terrestrial Planets, Workshop on Remote Sensing of Volcanic Gases, and a Workshop on Glass and Ceramics Industry in Space Based on Lunar Materials. Each of these conferences results in a variety of publications usually including an abstract volume prepared before the meeting, a proceedings volume which may be published as an independent monograph or as a special volume in a journal, or in a workshop summary which is distributed in the Institute's Contributions series.

A library to provide literature support to the visiting and staff scientists at the Institute and JSC was part of the original organization of the Institute. It contained a basic collection of geology and astronomy texts and journals and provided a full-range of library services to the community. In 1973, the collection of Apollo data consisting of the photography, the cartography, and supporting documents which was housed at the Johnson Space Center was transferred to the Institute and a Lunar Data Center including the Library, the Photo/Map Library, the Lunar Sample Information Library, and the Geophysical Data Center was established.

Although NSSDC remained the major distribution center for lunar data, it became apparent that the specialized and in-depth knowledge of the collections and workings of the data resulting from the missions, which the experienced staff of the Institute had acquired, would be useful to the scientist, educator, or student attempting to access the data.

For example, a knowledge of the history of the organization of the lunar sample collection is highly valuable to properly access the information available at the Johnson Space Center and at NSSDC.

The Lunar Samples are a data set which presented some unique and exciting challenges. The Curatorial Facility at the Johnson Space Center is responsible for the use and preservation of the samples.

Each sample has been carefully documented and a record of the allocation of the samples, the experiments performed, and the results of those experiments are filed in a data pack at this Facility. As the samples of each mission were returned to earth, a special team called the Preliminary Examination Team (PET) carefully opened each sample bag collected by the astronauts and did a visual and physical description of the sample. A series of photographs was taken in the laboratory to show all six faces of each sample. These macro descriptions done by members of the PET and at least one of

the photographs were published in a series of Lunar Sample Catalogs.

A method of cataloging the samples was devised. In general, a five digit generic number was assigned to each sample. The system of numbering evolved through the Program and became more significant in the later missions. The general scheme is:

Apollo 11	10abc
Apollo 12	12abc
Apollo 13	14abc
Apollo 15	15abc
Apollo 16	6xabc
Apollo 17	7xabc
Luna 16	21abc and/or 11016
Luna 20	22abc and/or 20abc
Luna 24	24 abc

For Apollo 16 and 17 the second digit, "x" is indicative of the station at which the sample was procured. For Apollo 15, 16, and 17, the last digit "c" is indicative of sample types such as 0=Unsieved Fines; 1=less than 1 mm fines; 2=1-2 mm fines; 3=2-4 mm fragments; 4= 4-10 mm fragments; and 5-9=rocks and greater than 1 cm fragments. A full description of the numbering system for the samples is contained in the Lunar Sample Information Catalogs for each mission. (5)

As each sample was divided and allocated to investigators for various kinds of analyses, a daughter sample was designated by a comma and a sequential number following the generic number. Thus 15455,54 is a section of Apollo 15 sample 15455. As the samples were divided, cutting diagrams were drawn so that the history of the sample would be documented.

The Lunar Sample Information Library at the Institute contains these catalogs, a mini data pack consisting of the sample photographs, description, cutting diagrams, and in some instances, photomicrographs, and many other documents which have been written on specific sample types by the various investigators working at the Curatorial Facility.

To maintain a record of the actual analyses which were obtained from the experiments conducted by the researchers, the Curator's Office maintains a "Lunar Sample Analysis Data Base." This data base contains the analytical data for each sample as reported in a selected set of published literature. The base is arranged by sample number and in the first section records each complete analysis plus the authors and bibliographic reference in which the analysis is published. The second part also arranged by sample number, lists the elemental analyses recorded for each sample along with the accession number of the bibliographic reference in which the analysis is reported. The third section is the bibliography arranged by accession number.

To obtain information about a particular sample, a researcher might consult the mini-data pack at the Institute, run a search for a particular element on the Sample Analysis Data Base, conduct a literature search through the Lunar and Planetary Bibliography maintained at the Institute, and then prepare a proposal for submission to the Lunar and Planetary Review Panel. Based on the results of his investigations of the data available, the researcher could then query NSSDC for the specific information he would need to support his research effort.

Another significant set of data which resulted from the scientific packages left on the lunar surface is called the Geophysical Data. Although this data is archived at the NSSDC, its usefulness to the researcher is enhanced by a review of what the data set contains.

The Apollo Lunar Surface Experiments Package was a completely self-contained science station deployed and activated by the Apollo astronauts and left on the lunar surface. The ALSEP collected scientific data on the lunar surface and transmitted the data to earth where the information was collected as part of the ALSEP support operations conducted at the Johnson Space Center. A forerunner of ALSEP, known as the Early Apollo Scientific Experiment Package (EASEP) was deployed by the Apollo 11 crew.

Experiments included in the package on various missions were: Passive Seismic Experiment, Active Seismic Experiment, Lunar Surface Magnetometer, Solar Wind Spectrometer, Suprathermal Ion Detector, Heat Flow, Charged Particle, Cold Cathode Gage, Lunar Ejecta and Meteorites, Lunar Seismic Profiling, Lunar Mass Spectrometer, Lunar Surface Gravimeter, and Dust Detector.

Scientific analysis of ALSEP data was accomplished by NASA contracts with specific investigators, and these contracts stipulated the archiving of analyzed data. To ensure proper data archiving, JSC management created the Geophysical Data Evaluation Working Group. The Group was asked to study the data processing and make recommendations on the archiving and distribution most appropriate for the present and future needs of the community. The data resulting from these experiments is collected and archived at NSSDC.

To make the data archived as accessible to the scientist as possible, the Institute was designated a "Geophysical Data Subcenter" by NASA, the collection of microfilm and microfiche data stored at NSSDC was duplicated and was incorporated, along with supporting documentation, into the Geophysical Data Center at the Institute. A researcher can review the data at the Institute, select what he needs and request just those subsets of information from the NSSDC.

Since much of the data archived is in the form of magnetic tapes, the Geophysical Data Curator at the Johnson Space Center undertook the project to prepare computer programs and associated user's guides which would make it possible to read these magnetic tapes. These programs and guides are available from the Geophysical Data Curator at the Johnson Space Center. (6,7)

The largest data set resulting from the various lunar exploration programs is the photography. This data set actually began with the Ranger series on July 31, 1964 and ended with Apollo 17 on December 19, 1972. The photography exists in many formats and media ... 16 mm, 35 mm, 70 mm, negative and positive transparencies, in color or black and white, stereoscopic close-ups and orbital panoramas. This data set is probably one of the most frequently used sets at NSSDC. However, to prepare a request for data, it is often most helpful to the user to visually review the photography in the various formats available.

The Photo/Map Library at the Institute has a complete collection of the Ranger, Orbiter, Surveyor, and Apollo photography of the Moon and various photographic indexes which provide access to the collection. It also has a selection of Zond and Luna photography. The collection is particularly user-oriented with the photographs arranged in notebooks, kardex files, or flip-files allowing the user to browse through the collection easily. Some selected sets of the photography have been assembled ranging from an approximate 100-item set useful to the educator or student, to the full-range of camera clicks taken on a particular mission. Personnel well-versed in the collection can assist the researcher to locate the items needed and then request them from NSSDC. If a researcher is unable to come to the Institute, the staff of the Photo/Map Library can respond to letter and telephone requests and make recommendations of those photos which would best assist the off-site researcher in his studies.

Another significant data set, which has been part of man's knowledge of the Moon since the days of Galileo, is the collection of lunar cartography. From the early hand-drawn efforts of Galileo, Hevelius, and Ricciolo through the 1:1,000,000 scale set of lunar astronautical charts (LAC) prepared by the U.S. Air Force Aeronautical Chart and Information Center (ACIC), man had been dependent on telescopic observation and photography for the details of his lunar maps. With the return of over 1,654 high-quality lunar photographs from the Lunar Orbiter flights in 1966-1967, man's 357-year dependence on the telescope for lunar mapping ended.

Several significant maps on various scales were prepared during the period between 1966 and 1972. The small scale maps (1:2,000,000 and smaller) have served for reference and planning purposes and have been used as the basis for many specialized maps prepared in support of the Apollo missions. Among these are:

- Lunar Planning Chart (LOC) series. 1:2,750,000. 1969-1971
- Lunar Equatorial Zone Mosaics. 1:2,500,000. 1968-1969
- Lunar Earthside, Farside and Polar Charts (LMP). 1:5,000,000. 1970 (new edition issued 1979)
- USAF Lunar Reference Mosaic (LEM-1 and LEM-1A). 1:5,000,000 and 1:10,000,000. 1966-1967
- Lunar Earthside Maps, Topographic. Scale varies. 1962-1964

In the medium scale (1:250,001 - 1:1,999,999) the following maps are representative:

- Lunar Astronautical Chart (LAC) series. 1:1,000,000. 1965-1967. (42 sheets)
- Apollo Intermediate Chart (AIC). 1:500,000. 1965-1967. (20 sheets)

Production of large scale lunar maps (1:250,000 and larger) became possible with the availability of spacecraft photography at scales allowing detailed description of lunar features. Series of maps were prepared from Orbiter, Surveyor, and Ranger photography. These maps were used in the studies to select the Apollo landing sites.

In 1973, NASA in conjunction with the Defense Mapping Agency (DMA) commenced a mapping program that would lead to the production of a series of 1:250,000 maps based on the photography from the Apollo 15, 16, and 17 missions. This program has resulted in the production of orthophotomaps and topographic orthophotomaps. A number of special scale maps have been prepared including 1:1,000,000 and 1:5,000,000. Detailed maps of the Apollo landing sites, called Traverse Charts, show the routes taken by the astronauts on the lunar surface. A new program to revise some of the LAC charts is currently underway. (8,9,10)

In addition to the lunar maps and charts based on telescopic and spacecraft photography, the U.S. Geological Survey (USGS) has prepared a series of maps at the scale of 1:1,000,000 showing the geology of most of the visible side of the Moon. Other geological maps on the Moon's hemispheres have been published in the 1:5,000,000 scale. These maps in the Survey's Miscellaneous Investigations Map Series (I-No.) are available from the USGS Publications Office.

A full complement of the lunar maps and charts including many of the earlier maps and atlases, and some of the Russian maps and charts, are available for study and in some cases, for loan, in the Photo/Map Library at the Institute. In many instances, staff personnel can assist the researcher to obtain copies of those maps needed. At present, most of the maps are available through the Planetary Programs Office at NASA Headquarters. However, plans are being formulated to make the maps available through the NSSDC.

As the emphasis of the space program shifted from the Moon to the other planets, the scope of the Institute also changed. In 1978, the Lunar Science Institute became the Lunar and Planetary Institute (LPI). The Lunar Data Center was reorganized to include:

- The Library/Information Center
- The Laboratory for the Analysis of Planetary Surfaces (LAPS) which includes:
  - The Photo/Map Library
  - The Geophysical Data Facility
  - The Lunar Sample Information Library
  - Photogeologic Interpretation Facility

And just as the sets of data resulting from Mariner, Viking, Voyager, and other planetary missions are being archived at NSSDC, the LPI is expanding its collections and information services to assist the lunar and planetary community.

The scope of the Bibliography of Lunar Literature maintained by the Library/Information Center was expanded in 1978 to include the Moon, planets, asteroids, comets, and meteorites. Indexing using a controlled thesaurus was also begun with the 1978 literature to increase the access points from just author and title key-words to a full range of subject-oriented index terms. Currently the on-line bibliography contains about 17,000 references. Searches are run based on telephone, letter, or in-person requests.

The Lunar and Planetary Information Bulletin (formerly the Lunar Science Information Bulletin) is compiled and published quarterly by the staff of the Information Center. It is circulated free to approximately 3200 scientists, educators, and students. The Bulletin contains news about personnel changes, conferences, publications, reports and reviews of space missions and mission opportunities, a calendar of events, and the listing of the current bibliography.

The collections of the Photo/Map Library have expanded to include Mariner, Viking, and Voyager imagery and maps. The present planetary imagery collection consists of:

Mariner 6	8x10 prints
Mariner 7	8x10 prints
Mariner 9	microfiche, 8x10 prints
Mariner 10	70mm negatives, positives, contact prints
Viking Landers-Orbiters	
	5" prints and negatives
	8x10 prints and negatives, mosaics
	20x24 prints, mosaics

Viking Orbiter 1

8x10 color mosaics
slides, color mosaics

Voyager I
Jupiter

Selected frames; films of mission overview, Jupiter rotation and motion of the Great Red Spot

Satellites

5" prints, negatives and positives
10" prints

A lunar and planetary slide collection is accessed by a computerized subject index. Slides may be borrowed from this collection to support lectures, publications, or other educational and research efforts.

The map collection is also expanding to include the planetary maps for Mercury and Mars prepared by USGS using Mariner imagery. These maps are available as shaded relief, topographic, and geologic representations. Additional planetary maps prepared from Viking and Voyager photography are being considered by the Mapping Program of the Planetary Programs Office at NASA. These maps will be included in the collection as they become available.

Through the development of LAPS, equipment for data analysis at the Institute includes a plane-table digitizer, Zeiss TGZ-3 Particle-Size Analyzer, Traid Viewer, Itek Variable Magnifying Viewer, Zoom Transferscope, and stereo viewers. Proposals for research projects using the LAPS facility and the collections of the Photo/Map Library may be submitted to the Director of the LPI.

To assist users to identify the lunar and planetary data which is available, the NSSDC and the LPI prepare catalogs and information resource guides. These are available from the issuing agency. (11,12,13)

Thus although the NSSDC will continue to be the major international distribution center of lunar and planetary data, the LPI and its staff serves as a liaison between the scientific and lay community and the data sets themselves.

REFERENCES

(1) International Council of Scientific Unions. Panel on World Data Centres, Third Consolidated Guide to International Data Exchange through the World Data Centres, Washington, ICSU Panel, 1973, 72 pp.

(2) WDC-A; Wold Data Center-A for Glaciology, Meteorology, Oceanography, Rockets and Satellites, Rotation of the Earth, Solar-Terrestrial Physics, Solid-Earth Geophysics, Washington, Environmental Data Service, 1978, 16 pp.

(3) Report on Rocket and Satellite Information and Data Exchange, Greenbelt, Md., World Data Center A, April 20, 1979, 14 pp.

(4) Vostreys, R.W., Planetary Data at the National Space Science Data Center, IN: Reports of Planetary Geology Program, 1978-1979, edited by J. Boyce and P.S. Collins, NASA TM-80339, 1979, p. 444.

(5) Warner, J. and C. Dardano, User Guide to the Lunar Sample Data Base, Houston, TX, Johnson Space Center, Curator's Office, July 3, 1978, 11 pp.

(6) Lauderdale, W.W. and W.F. Eichelman, The Apollo Scientific Data Experiments Handbook, Houston, TX, Johnson Space Center, NASA TM-X-58131, 1974, 898 pp.

(7) Bates, J.R., W.W. Lauderdale and H. Kernaghan, ALSEP Termination Report, Washington, D.C., NASA, NASA RP-1036, April 1979, 162 pp.

(8) Schimerman, L.A., Lunar Cartographic Dossier, v.1, St. Louis, MO, Defense Mapping Agency, Aerospace Center, 1975, looseleaf.

(9) Kinsler, D.K., User Guide to 1:250,000 Scale Lunar Maps, Houston, TX, LSI, 1975, 24 pp. (LSI Contribution No. 206)

(10) Kinsler, D.C., New Lunar Cartographic Products, Proc. Lunar Science Conf. 7th v.3, New York, Pergamon Press, 1976, p. i-x.

(11) Cameron, W.S., E.J. Mantell and E.R. Miller, Catalog of Lunar Mission Data, NSSDC 77-02, 1977, 204 pp.

(12) Vostreys, R.W., Catalog of Viking Mission Data, NSSDC 78-01, 1978, 36 pp.

(13) Waranius, F.B. and M.A. Hager, Lunar and Planetary Science Information Resources, Houston, LPI, 1976, 24 pp.

ACKNOWLEDGMENT. The author wishes to thank R.W. Vostreys, NSSDC, and Ron Weber, Dan Kinsler, and K. Leung, LPI for their help in preparing this paper. The Lunar and Planetary Institute is operated by the Universities Space Research Association under Contract No. NSR-09-051-001 with the National Aeronautics and Space Administration. This paper is Lunar and Planetary Institute Contribution No. 382.

USEFUL ADDRESSES

World Data Center A
Coordination Office
National Academy of Sciences
2101 Constitution Avenue
Washington, DC 20418

National Space Science Data Center
Code 601.4
Goddard Space Flight Center
Greenbelt, Md 20771

World Data Center A
Rockets and Satellites
Code 601
Goddard Space Flight Center
Greenbelt, MD 20771

Curator's Office
NASA Johnson Space Center
Code: SN2
Houston, TX 77058

Geophysical Data Curator
NASA Johnson Space Center
Code: SN6
Houston, TX 77058

Planetary Programs Office
NASA Headquarters
Code: SL-4
Washington, DC 20546

AGARD

NATO OTAN

7 RUE ANCELLE · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone 745.08.10 · Telex 610176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD – VSL
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, 1140 Bruxelles

CANADA
Defence Scientific Information Service
Department of National Defence
Ottawa, Ontario K1A OZ2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92 Châtillon sous Bagneux

GERMANY
Zentralstelle für Luft- und Raumfahrt-
dokumentation und -information
c/o Fachinformationszentrum Energie,
Physik, Mathematik GmbH
Kernforschungszentrum
7514 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force General Staff
Research and Development Directorate
Holargos, Athens, Greece

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direcção do Serviço de Material
da Forca Aerea
Rua da Escola Politecnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA)
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy
National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22161, USA

Microfiche
Space Documentation Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

Microfiche
Technology Reports
Centre (DTI)
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RF
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Announcements (GRA)
published by the National Technical
Information Services, Springfield
Virginia 22161, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London W1P 1HD